

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-222398

(43)Date of publication of application : 11.08.2000

(51)Int.Cl.

G06F 17/21
G06F 13/00
G06F 17/30

(21)Application number : 2000-020499

(71)Applicant : FUJI XEROX CO LTD

(22)Date of filing : 28.01.2000

(72)Inventor : BICKMORE TIMOTHY W
SCHILIT WILLIAM N
ANDREAS GARGENSHORN
SULLIVAN JOSEPH W

(30)Priority

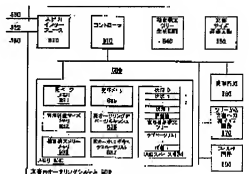
Priority number : 99 239294 Priority date : 29.01.1999 Priority country : US

(54) METHOD AND SYSTEM FOR FILTERING DOCUMENT

(57)Abstract:

PROBLEM TO BE SOLVED: To display a document which is designed for a large display area in a small display area by acquiring a request for a document from a decentralized network and putting the requested document back into a reauthoring-completed version.

SOLUTION: A filter memory 690 inputs a requested filter from a user through a communication link 522 or 560 under the control of a controller 610. The requested filter is supplied from a node of the decentralized network through a communication link 580. The filter memory 690 inputs the requested document from the node, and filters out and extracts the elements of a requested page from the original document by using an abstract syntax tree generated by an abstract syntax tree generation memory 633. The filter memory 690 stores the extracted page elements in an original page memory 631 instead of the original document. Then an document authoring system 600 combines the



Claim(s)

[Claim 1] The step which is the approach of filtering said document in order to extract the desirable element of a document, and offers a predetermined filter, The step which obtains said document which should be filtered, and the step which generates an analysis tree from said obtained document, The step which identifies at least one element of said document which should be navigated and extracted through said analysis tree based on said offered filter, The filtering approach of a document equipped with the step which extracts said at least one identified element, and the step which forms the version by which said document was filtered combining said at least one extracted element.

[Claim 2] The step which offers said predetermined filter is a method according to claim 1 of having the step which searches said predetermined filter from the node of a distributed network.

[Claim 3] The step which offers said predetermined filter is an approach [equipped with the step which offers the uniform resource locator (URL) to said predetermined filter to the host node of a distributed network, and the step which searches said predetermined filter from the node of said distributed network corresponding to said offered URL] according to claim 1.

[Claim 4] The step which offers URL to said predetermined filter is an approach [equipped with the step which transmits said URL from the data processor connected to said distributed network] according to claim 3.

[Claim 5] The step which transmits said uniform resource locator from said data processor is an approach [equipped with the step which transmits said URL through a radio channel] according to claim 4.

[Claim 6] The step which searches said predetermined filter from said node of said distributed network is an approach [equipped with the step which searches said predetermined filter from said host node of said distributed network] according to claim 3.

[Claim 7] The step which searches said predetermined filter from said node of said distributed network is an approach [equipped with the step which searches said predetermined filter from the remote node of said distributed network] according to claim 3.

[Claim 8] The step which obtains said said document which should be filtered is an approach [equipped with the step which searches said document which should be filtered from the node of a distributed network] according to claim 1.

[Claim 9] The step which obtains said document which should be filtered is an approach [equipped with the step which offers URL to said document which should be filtered, and the step which searches said document which should be filtered from the node of a distributed network corresponding to said offered URL] according to claim 1.

[Claim 10] The step which offers said URL to said document which should be filtered is an approach [equipped with the step which embeds said URL to said document which should be filtered in said predetermined filter] according to claim 9.

[Claim 11] The step navigated through said analysis tree is an approach according to claim 1 containing at least one of the steps which face a context current within the step which moves to the current context of said analysis tree further, the step which moves outside from the current context of said analysis tree, and said analysis tree, and move to other structures of said analysis tree.

[Claim 12] For said current context, the step which is the current node of said analysis tree and moves to the current context of said analysis tree further is an approach [equipped with the step which moves to at least one child node of a current node to said current node of said analysis tree] according to claim 11.

[Claim 13] The step which moves to the current context of said analysis tree equipped with the step which moves to at least one child node of a current node to said current node of said analysis tree further is the approach according to claim 12 have the step which uses GO "regular representation" command, and a child node at least contains said string defined by aforementioned "regular representation."

[Claim 14] For said current context, the step which is the current node of said analysis tree and moves outside from said current context of said analysis tree is an approach [equipped with the step which moves to at least one parent node of a current node from the current node of said analysis tree]

according to claim 11.

[Claim 15] The step which said current context is the current node of said analysis tree, and moves to the next structure of said analysis tree to a current context within said analysis tree is an approach [equipped with the step which moves through the node of said analysis tree from said current node] according to claim 11.

[Claim 16] Said step which moves through the node of said analysis tree from said current node is an approach [equipped with the step which moves sequentially through said analysis tree from said current node] according to claim 15.

[Claim 17] Said step navigated through said analysis tree is an approach according to claim 11 further equipped with any of migration in the current context, the migration from a current context to outside, and the migration that faces a current context they are based on matching with the regular representation defined by said searched filter.

[Claim 18] Said step navigated through said analysis tree is an approach according to claim 11 further equipped with any of migration in the current context, the migration from a current context to outside, and the migration that faces a current context they are based on repetitive matching with the regular representation defined by said searched filter.

[Claim 19] Said step navigated through said analysis tree is an approach according to claim 11 further equipped with any of migration in the current context, the migration from a current context to outside, and the migration that faces a current context they are based on repetitive activation of the navigation command defined by said searched filter.

[Claim 20] Said step navigated through said analysis tree is an approach [equipped with the step which moves to at least one desirable location within said analysis tree] according to claim 1 based on describing structurally each ** better ***** in said analysis tree.

[Claim 21] Said step navigated through said analysis tree is an approach [equipped with the step which moves to at least one desirable location within said analysis tree] according to claim 1 based on describing the regular representation included in the context of each ** better *****.

[Claim 22] Furthermore, the step which generates the document containing said said at least one extracted element, The step which generates an evaluation value from said generated document, and the step which determines whether said evaluation value satisfies at least one valuation basis, The step which outputs said document when satisfying the valuation basis said whose evaluation value is said one, An approach [equipped with the step which transforms said document into a deformed document and repeats said step to generate and said step to determine when not satisfying the valuation basis said whose evaluation value is said one] according to claim 1.

[Claim 23] The document filtering system which is a document filtering system which filters said document automatically based on the selected filter in order to extract at least one element from a document, and is equipped with analysis tree generation memory, document filtering memory, and the extracted element output memory.

[Claim 24] Said analysis tree generation memory is a document filtering system according to claim 23 which analyzes said document in order to generate an analysis tree.

[Claim 25] Said document filtering system is a document filtering system according to claim 24 navigated within said analysis tree based on the navigation command of said selected filter.

[Claim 26] Said document filtering system is a document filtering system according to claim 24 which extracts a document element from said document based on the report command of said selected filter, and the current context of said analysis tree.

[Claim 27] Said document filtering system is a document filtering system according to claim 26 which determines the current context of said analysis tree by navigating within said analysis tree based on the navigation command of said selected filter.

[Claim 28] Furthermore, a document filtering system [equipped with document size evaluation memory and deformation memory] according to claim 23.

[Claim 29] Said document size evaluation memory is a document filtering system according to claim 28 which evaluates the analysis tree generated by said analysis tree generation memory from the filtered

document containing said at least one extracted element, and determines whether said document satisfies at least one valuation basis.

(19) 日本特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-222398

(P2000-222398A)

(43) 公開日 平成12年8月11日 (2000. 8. 11)

(51) Int.Cl. ⁷	識別記号	F I	ページコード* (参考)
G 0 6 F 17/21		C 0 6 F 15/20	5 8 0 L
13/00	3 5 4	13/00	3 5 4 D
17/30		15/40	3 4 0
			3 7 0 A
審査請求 未請求 請求項の数29 O L (全 34 頁)			
(21) 出願番号 特願2000-20499(P2000-20499)	(71) 出願人 000003496		
(22) 出願日 平成12年1月28日 (2000. 1. 28)	富士ゼロックス株式会社		
(31) 優先権主張番号 2 3 9 2 9 4	東京都港区赤坂二丁目17番22号		
(32) 優先日 平成11年1月29日 (1999. 1. 29)	(72) 発明者 ティモシー ダブリュー、 ビックモア		
(33) 優先権主張国 米国 (U S)	アメリカ合衆国 02144 マサチューセツ		
	州 サマービル ライト ホール スト		
	リート 13		
	(72) 発明者 ウィリアム エヌ、 シリット		
	アメリカ合衆国 94025 カリフォルニア		
	州 メンロ パーク メンロ アベニュー		
	973		
	(74) 代理人 100079049		
	弁理士 中島 淳 (外 1 名)		

最終頁に続く

(54) 【発明の名称】 文書のフィルタリング方法及び文書フィルタリングシステム

(57) 【要約】

【課題】 デスクトップコンピュータスクリーン上の表示のために元来デザインされた文書をより小さな表示スクリーンでの表示のために再オーサリングすること。

【解決手段】 所定のフィルタを提供するステップと、フィルタリングされるべき前記文書を得るステップと、前記得られた文書から解析ツリーを生成するステップと、前記識別されたフィルタに基づいて前記解析ツリーを介してナビゲートして抽出されるべき前記文書の少なくとも一つのエレメントを識別するステップと、前記少なくとも一つの識別されたエレメントを抽出するステップと、前記少なくとも一つの抽出されたエレメントを組合せて前記文書のフィルタリングされたバージョンを形成するステップと、を備える。

【特許請求の範囲】

【請求項1】 文書の望ましいエレメントを抽出するために前記文書をフィルタリングする方法であって、所定のフィルタを提供するステップと、フィルタリングされるべき前記文書を得るステップと、前記得られた文書から解析ツリーを生成するステップと、前記提供されたフィルタに基づいて前記解析ツリーを介してナビゲートして抽出されるべき前記文書の少なくとも一つのエレメントを識別するステップと、前記少なくとも一つの識別されたエレメントを抽出するステップと、前記少なくとも一つの抽出されたエレメントを組合せて前記文書のフィルタリングされたバージョンを形成するステップと、を備える文書のフィルタリング方法。

【請求項2】 前記所定のフィルタを提供するステップは、分散形ネットワークのノードから前記所定のフィルタを検索するステップを有する、請求項1に記載の方法。

【請求項3】 前記所定のフィルタを提供するステップは、前記所定のフィルタに対するユニフォームリソースロケータ（URL）を分散形ネットワークのホストノードへ提供するステップと、前記提供されたURLに対応して、前記分散形ネットワークのノードから前記所定のフィルタを検索するステップと、を備える、請求項1に記載の方法。

【請求項4】 前記所定のフィルタに対するURLを提供するステップは、前記URLを前記分散形ネットワークに接続されたデータ処理装置から送信するステップを備える、請求項3に記載の方法。

【請求項5】 前記データ処理装置から前記ユニフォームリソースロケータを送信するステップは、前記URLを無線通信チャネルを介して送信するステップを備える、請求項4に記載の方法。

【請求項6】 前記分散形ネットワークの前記ノードから前記所定のフィルタを検索するステップは、前記分散形ネットワークの前記ホストノードから前記所定のフィルタを検索するステップを備える、請求項3に記載の方法。

【請求項7】 前記分散形ネットワークの前記ノードから前記所定のフィルタを検索するステップは、前記分散形ネットワークの遠隔ノードから前記所定のフィルタを検索するステップを備える、請求項3に記載の方法。

【請求項8】 前記フィルタリングされるべき前記文書を得るステップは、分散形ネットワークのノードから前記フィルタリングされるべき文書を検索するステップを備える、請求項1に記載の方法。

【請求項9】 フィルタリングされるべき前記文書を得るステップは、前記フィルタリングされるべき文書に対するURLを提供するステップと、前記提供されたURLに対応して、分散形ネットワークのノードから前記フィルタリングされるべき文書を検索するステップと、を備える、請求項1に記載の方法。

【請求項10】 前記フィルタリングされるべき文書に対する前記URLを提供するステップは、前記フィルタリングされるべき文書に対する前記URLを前記所定のフィルタに埋め込むステップを備える、請求項9に記載の方法。

【請求項11】 前記解析ツリーを介してナビゲートするステップは、前記解析ツリーの現在のコンテキストに更に移動するステップと前記解析ツリーの現在のコンテキストから外に移動するステップと前記解析ツリー内で現在のコンテキストに相対して前記解析ツリーの他の構造へ移動するステップの内の少なくとも一つを含む、請求項1に記載の方法。

【請求項12】 前記現在のコンテキストは、前記解析ツリーの現在のノードであり、前記解析ツリーの現在のコンテキストに更に移動するステップは、前記解析ツリーの現在のノードから前記現在のノードの少なくとも一つの子ノードへ移動するステップを備える、請求項11に記載の方法。

【請求項13】 前記解析ツリーの現在のノードから前記現在のノードの少なくとも一つの子ノードへ移動するステップを備える前記解析ツリーの現在のコンテキストに更に移動するステップは、GO正則表現コマンドを使用するステップを備え、前記少なくとも一つの子ノードは前記正則表現によって定義されるストリングを含む、請求項12に記載の方法。

【請求項14】 前記現在のコンテキストは、前記解析ツリーの現在のノードであり、前記解析ツリーの前記現在のコンテキストから外へ移動するステップは、前記解析ツリーの現在のノードから現在のノードの少なくとも一つの親ノードへ移動するステップを備える、請求項11に記載の方法。

【請求項15】 前記現在のコンテキストは、前記解析ツリーの現在のノードであり、前記解析ツリー内で現在のコンテキストに対して前記解析ツリーの次の構造へ移動するステップは、前記現在のノードから前記解析ツリーのノードを介して移動するステップを備える、請求項11に記載の方法。

【請求項16】 前記現在のノードから前記解析ツリーのノードを介して移動する前記ステップは、前記現在のノードから前記解析ツリーを介してシーケンシャルに移動するステップを備える、請求項15に記載の方法。

【請求項17】 前記解析ツリーを介してナビゲートす

る前記ステップは、前記検索されたフィルタで定義された正則表現とのマッチングに基づいて、現在のコンテキストへの移動、現在のコンテキストから外への移動、及び現在のコンテキストに相対する移動の内の何れかを更に備える、請求項11に記載の方法。

【請求項18】 前記解析ツリーを介してナビゲートする前記ステップは、前記検索されたフィルタで定義された正則表現との反復マッチングに基づいて、現在のコンテキストへの移動、現在のコンテキストから外への移動、及び現在のコンテキストに相対する移動の内の何れかを更に備える、請求項11に記載の方法。

【請求項19】 前記解析ツリーを介してナビゲートする前記ステップは、前記検索されたフィルタで定義されたナビゲーションコマンドの反復実行に基づいて、現在のコンテキストへの移動、現在のコンテキストから外への移動、及び現在のコンテキストに相対する移動の内の何れかを更に備える、請求項11に記載の方法。

【請求項20】 前記解析ツリーを介してナビゲートする前記ステップは、前記解析ツリー内の各望ましい位置を構造的に記述することに基づいて、前記解析ツリー内の少なくとも一つの望ましい位置へ移動するステップを備える、請求項1に記載の方法。

【請求項21】 前記解析ツリーを介してナビゲートする前記ステップは、各望ましい位置のコンテキスト内に含まれる正則表現を記述することに基づいて、前記解析ツリー内で少なくとも一つの望ましい位置へ移動するステップを備える、請求項1に記載の方法。

【請求項22】 更に、前記少なくとも一つの前記抽出された要素を含む文書を生成するステップと、前記生成された文書から評価値を生成するステップと、前記評価値が少なくとも一つの評価基準を満足するか否かを決定するステップと、

前記評価値が前記一つの評価基準を満足する場合、前記文書を出力するステップと、前記評価値が前記一つの評価基準を満足しない場合、前記文書を変形済文書へ変形して、前記生成するステップ及び前記決定するステップを繰り返すステップと、を備える請求項1に記載の方法。

【請求項23】 文書から少なくとも一つの要素を抽出するために、選択されたフィルタに基づいて、前記文書を自動的にフィルタリングする文書フィルタリングシステムであって、解析ツリー生成メモリと、文書フィルタリングメモリと、抽出された要素出力メモリと、を備える文書フィルタリングシステム。

【請求項24】 前記解析ツリー生成メモリは、解析ツリーを生成するために前記文書を解析する、請求項23に記載の文書フィルタリングシステム。

【請求項25】 前記文書フィルタリングシステムは、

前記選択されたフィルタのナビゲーションコマンドに基づいて、前記解析ツリー内でナビゲートする、請求項24に記載の文書フィルタリングシステム。

【請求項26】 前記文書フィルタリングシステムは、前記選択されたフィルタのレポートコマンドと前記解析ツリーの現在のコンテキストに基づいて、前記文書から文書要素を抽出する、請求項24に記載の文書フィルタリングシステム。

【請求項27】 前記文書フィルタリングシステムは、前記選択されたフィルタのナビゲーションコマンドに基づいて、前記解析ツリー内でナビゲートすることによって、前記解析ツリーの現在のコンテキストを決定する、請求項26に記載の文書フィルタリングシステム。

【請求項28】 更に、文書サイズ評価メモリと、変形メモリと、を備える、請求項23に記載の文書フィルタリングシステム。

【請求項29】 前記文書サイズ評価メモリは、前記少なくとも一つの抽出された要素を含むフィルタリングされた文書から前記解析ツリー生成メモリによって生成された解析ツリーを評価して、前記文書が少なくとも一つの評価基準を満足するか否かを決定する、請求項28に記載の文書フィルタリングシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ワールドワイドウェブからの任意の文書を自動的に再オーサリングしてパーソナルデジタルアシスタント(PDA)及びセルラ電話のような小さなスクリーン装置上に適切に前記文書を表示する文書再オーサリングシステム及び方法に関し、これらのシステム及び方法は、ウェブへの装置独立アクセスを提供する。

【0002】

【従来の技術】パーソナル電子装置からワールドワイドウェブ(WWW)へのアクセスは、1995年3月のカリフォルニア州、サンフランシスコにおける、J. Barlettの"Experience with a Wireless World Web Client", IEEE COMPCON95、1994年10月のイリノイ州、シカゴにおける、第2回国際ワールドワイドウェブ会議でのS. Gessler等の"PDAs Mobile WWW Browsers", 1994年12月のカリフォルニア州、サンタクルーズにおける、移動体コンピューティングシステム及び応用(Mobile Computing System and Application)の研修会でのG. Voelker等の"Mobisac: An Information System for a Mobile Wireless Computing Environment", 及び1994年8月のT. Watsonの1994年移動体コンピューティングシステム及び応用の研究会討議資料の"Application Design for Wireless Computing"のような研究プロジェクトで論証されている。このようなアクセスは、現在商業的に現実の物となっている。General M

agicのSony's MagicLinkに対するPresto! Links及びA11PenのNewtonに対するNetHopper及びShrapのMI-10の全てがPDAクラス装置に対するWWWブラウザを提供すると共に、Nokia 9000Communicator及びSamsungのDuettは、セルラー電話からのウェブアクセス能力を有している。

【0003】残念なことに、ワールドワイドウェブ及び他の分散形ネットワーク上の多くの文書は、少なくとも640x480の解像度のカラーモニターを有するデスクトップコンピュータで表示するように設計されている。多くのページは、一層大きな解像度を持つよう意図されて設計されている。反対に、多くのPDA（個人用携帯情報端末）クラス装置及びセルラー電話表示は、一層小さくなっている。この表示領域における差によって設計表示領域対利用可能表示領域の比が4対1から10対1若しくはそれを超え、それによってこれらの小さな装置上での大部分のワールドワイドウェブ文書の直接表示が不快感を与え、共にナビゲートできなくなり、最悪の場合、判読できなくなる。これは、これらの小さな装置を使用しているワールドワイドウェブページへのアクセスにおける主要な問題を提起する。即ち、HTML（ハイパーテキスト記述言語）文書のような、デスクトップシステム用に設計された任意のウェブ文書を大きく限られた表示能力しかないパーソナル電子装置上にどのように表示するかという問題である。

【0004】このような従来の技術は、コンピュータの移動性と無線での接続性を提供するが、小さなスクリーン上で文書及びウェブページを見るための標準的解決策は、ユーザが拡大鏡を持ち運ぶ場合は重要である、スクリーン解像度を上げる、またはファックス又はプリントする能力をローカルハードコピー装置へ提供することである。しかし、それらは、共に不便であり、最初の場所で電子文書を有することの合理性と矛盾する。小さなスクリーン装置上にウェブ文書を表示するための五つの一般的なアプローチがある。それらのアプローチは、装置指定オーサリング、複数装置オーサリング、クライアントサイドナビゲーション、自動再オーサリング、及びページフィルタリングである。装置指定オーサリングは、例えば、Nokia 9000のような表示及び通信ソフトウェアが外付けされたセルラー電話のような特定の表示装置のために一セットのウェブ文書をオーサリングすることを含む。このアプローチの基本的考え方は、そのような特定の装置のユーザが選択された一セットのサービスに対するアクセスを有するのみであることである。このように、これらのサービスのための文書は、アクセス装置の特定の表示システムのためにアップフロント（upfront）で設計されなければならない。情報は、大量に分散形ネットワークから提供され得るが、望ましいページが予め定義されなければならない、且つカスタム情報抽出及びページフォーマットソフトウェアが小さな装

置へその情報を受け渡すために書き込まなければならない。これは、Unwired Planet's UP. Linkサービスで取られるアプローチであり、それは、所有権を主張できるマークアップ言語（Markup Language）を使用する。

【0005】複数装置オーサリングにおいて、一つの範囲内の装置が識別される。次に、単一のソース文書から一セットのレンダリングされた文書へのマッピングが定義されて、識別された範囲内でそれらの装置をカバーする。この一例は、1995年11月のセント大学（University of Kent）のカンタベリーコンピュータリングラボラトリWWWページ（Canterbury Computing Laboratory WWW Page）でのI. Cooper等のPDA Web Browsers: Implementation Issuesで論じられているStretchTextアプローチである。StretchTextにおいて、ワードレベルにされる可能性のある文書の部分は、“抽象のレベル” 測度でタグ付けされ得る。文書を受信すると、ユーザは、それらのユーザが見ることを望む抽象のレベルを指定でき、対応する詳細又は詳細の欠如をもって提示される。

【0006】複数装置オーサリングの他の例は、HTMLカスケディングスタイルシート（CSS）であり、それは、1996年9月のワールドワイドウェブ協会（WWW Consortium）のH. Lie等の"Cascading Style Sheets"で述べられている。このカスケードスタイルシートにおいて、一つの単一シートは、文書の異なる構造部分に対して一セットの表示属性を定義する。例えば、全てのトップレベルセクションのヘッダーが、赤の18ポイントタイムフォントで表示されるように定義され得る。一連のスタイルシートは、文書に取り付けられることが出来、各スタイルシートは、その文書のオーサラーに対するこのスタイルシートの望ましさを述べる重みを備える。また、ユーザは、デフォルトスタイルシートを指定できる。また、分散形ネットワークアクセスするためにユーザによって使用されるブラウザは、“デフォルト”スタイルシートを定義できる。オーサラーのスタイルシートは、通常、ユーザのスタイルシートをオーバーライドする（無効にする）が、ユーザにユーザの特定の表示への文書のレンダリングを適合させる能力を提供することによって、ユーザは、選択的にオーサラーのスタイルシートをイネール又はディスイネールできる。

【0007】クライアントサイドナビゲーションにおいて、ユーザは、任意の所与の時に表示される単一のウェブページの部分を変更することによって、その単一のウェブページ内で対話的にナビゲートする能力が与えられる。この非常に些細な例は、文書表示リアでスクロールバーを使用することである。非常に複雑なアプローチは、1994年のACM PRESS（ACMプレス）のACM UIST '94の会報、B. Bederson等による"Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics"で述べられているように、PAD++シ

システムで取られているアプローチである。そのアプローチにおいて、ユーザは、無限の解像度で文書上で装置表示を自由にズーム及びパンすることが出来る。また、1994年10月のイリノイ州、シカゴにおける第2回国際ワールドワイドウェブ会議でのJ. Hue等の"Active Outlining for HTML Documents: An X-Mosaic Implementation"で述べられているように、アクティブアウトライン化は、クライアントサイドナビゲーション技術として実行されている。ここにおいて、ユーザは、夫々のセッションヘディング下で、文書のセクションをダイナミックに拡大及び崩壊できる。このカテゴリに入る他の技術は、1996年4月のカナダ国BC、バンクーバーでのComputer-Human Interaction: CHI'96の会報のT. Kamba等の"Using small screen space more efficiently"で述べられているような、半透明仕掛け(widgets)、及び1993年のSIGGRAPH'93 Conference ProceedingsのE.

Bier等の"Toolglass and Magic Lenses: The See-through Interface"で述べられているようなMagic Lenses (マジックレンズ) システムを含む。

【0008】自動文書再オーサリングは、ターゲット表示装置の特性と共に、デスクトップサイズのモニターで表示されるように設計された、HTML文書のような任意の文書を扱い、一連の変形を介してその任意の文書を再オーサリング出来るソフトウェアを含み、それによって、その任意の文書は、ターゲット表示装置上で適切に表示される。このプロセスは、クライアントによって、サーバーによって、又はこれらの変形サービスを提供するために単独で存在する。HTTP (ハイパーテキスト転送プロトコル) プロキシサーバーのような中間プロキシサーバーによって実行される。この後者のアプローチの例は、UC Berkeley Pythia プロキシサーバーであり、それは、1996年5月のフランス、パリでの第5回国際ワールドワイドウェブ会議のA. Fox等の"Reducing WWW Latency and Bandwidth Requirements by Real-Time Distillation"で述べられている。これは、ウェブページ画像の変形を実行する。しかしながら、Pythia プロキシサーバーの目的は、単独でページ検索時間を最小にするためである。Spyglass Prismは、ページタグ又は埋め込みオブジェクトタイプに関連する一定の変形を使用して、HTML文書の自動再オーサリングを実行する市販の製品である。例えば、Prismは、50%全てのJPEG画像を縮小する。

【0009】最後に、ウェブページフィルタリングは、ユーザにユーザが興味のあるページの部分のみを見させる。フィルタリングは、HTTPプロキシサーバーのような中間サーバーで実行されることができ、無線バンド幅及び装置メモリを節約する。しかしながら、フィルタリングは、表示管理技術のようなクライアント装置によっても実行される。フィルタの仕様は、キーワード又は正規表現マッチング、又はページ構造ナビゲーションと抽

出コマンドに基づくことができる。フィルタリングは、可視ツールを使用して又はスクリプト言語を使用して指定される。

【0010】装置指定オーサリング、マルチプル(複数)装置オーサリング、クライアントサイドナビゲーション、自動再オーサリング、及びページフィルタリングの五つのアプローチの各々は、独自の利点及び欠点がある。装置指定オーサリングは、一般的に人間のデザイナーの直接関与のお陰で最もよく見える結果を生ずる。しかしながら、装置指定オーサリングは、ユーザのアクセスをその指定装置のためにオーサリングされた小さな選択セットの文書に制限する。マルチプル装置オーサリングは、装置指定オーサリングよりも文書当りの全体の労力が少なくてすむが、単一のデスクトッププラットフォームのための単一バージョンの文書を単純にオーサリングする場合に比較してまだかなり多くの手作業による設計作業が必要である。クライアントサイドナビゲーションは、優れたセットのビューイング (viewing) テクニックが開発できる時はうまく働く。しかしながら、クライアントサイドナビゲーションは、全体の文書が一度にクライアント装置へ転送されなければならない。それは、貴重な無線帯域幅及びメモリを消費する。更に、PAD+で取られた"ヒープホール"は、大きな文書に対しては非常に使用しにくいように思え、多くのウェブページが厳格なセクション/サブセクション構成を使用しないか、その構成を不正確に使用するので、アクティブアウトライニング技術は適用性に制限がある。

【0011】このように、自動再オーサリングが、情報を失うことなく、読み易く、ナビゲーション可能で、審美的に満足な再オーサリング済文書を生成するように行われるならば、自動再オーサリングは、広範囲に亘る装置からウェブ文書や他のウェブ内容への広範囲に亘るアクセスを提供する理想的なアプローチである。

【0012】

【発明が解決しようとする課題】本発明の目的は、より大きな表示領域のために設計された文書をより小さな表示領域上での表示のために自動的に再オーサリングするシステム及び方法を提供することである。

【0013】また、本発明の目的は、各サブ文書がより少ない表示領域を必要とする、複数のリンクされたサブ文書に文書を自動的に変形するシステム及び方法を提供することである。

【0014】また、本発明の目的は、複数のセットのリンクされたサブ文書を生成するために、複数の異なる変形を原稿(オリジナル文書)に自動的に適用するシステム及び方法を提供することである。

【0015】更に、本発明の目的は、追加のリンクされたサブ文書を生成するために、前述の複数の異なる変形を前記複数のセットのリンクされたサブ文書の少なくとも一つへ自動的に提供するシステム及び方法を提供する

ことである。

【0016】更にまた、本発明の目的は、主サブ文書の最良の一つを決定するために、各セットのリンクされたサブ文書の主サブ文書を解析するシステム及び方法を提供することである。

【0017】また、本発明の目的は、前述の最良の主サブ文書がより小さな表示領域に表示され得るか否かを決定して、否の場合、更に要求される表示領域に縮小するために、その主サブ文書へ更なる変形を適用するシステム及び方法を提供することである。

【0018】また、本発明の目的は、より小さな表示領域に表示可能な文書の望ましい部分を抽出するために、文書をフィルタリングするシステム及び方法を提供することである。

【0019】更に、本発明の目的は、予め定義されたスクリプトに基づいて、望ましい部分を抽出するために、文書をフィルタリングするシステム及び方法を提供することである。

【0020】また、本発明の目的は、望ましい部分を抽出するために文書をフィルタリングするのに使用できるスクリプトを生成するシステム及び方法を提供することである。

【0021】更に、本発明の目的は、望ましい部分を抽出するために文書をフィルタリングのためのスクリプトを書き込むことが出来るスクリプト言語を提供することである。

【0022】

【課題を解決するための手段】一例示の実施の形態において、本発明の文書再オーサリングシステム及び方法は、所与の表示サイズのための最良に見える文書を達成するために、ヒューリスティックプランニング技術を使用して、要求されたウェブページ及び一セットの構造的ページ変形をダイナミックに再オーサリングするHTTPプロキシで実施される。本発明のシステム及び方法に従う自動文書再オーサリングは、クライアントによって、サーバによって、又は、一例示の実施の形態において、これらの変形サーバを提供するために単独で存在する中間HTTPプロキシサーバによって、実行され得る。更に、本発明に従う自動文書再オーサリングシステム及び方法は、これらの装置の組合せで実行され得る。

【0023】本発明の自動再オーサリングシステム及び方法は、PDAで見られる表示で良好に働く。しかしながら、本発明の文書再オーサリングシステム及び方法は、時々ナビゲートが困難なページを生成する。セルラ電話から、例えば、インターネット又はイントラネットのような、分散形ネットワークにアクセスする時、大部分のユーザは、主に非常に特定された情報に興味がある。本発明の文書フィルタリングシステム及び方法は、

このようなユーザに彼等/彼女らが表示されたい情報を定義するためにマニュアル制御を提供する。本発明の文書フィルタリングシステム及び方法は、容易にナビゲーション出来るページの小さな部分のみを戻す。本発明の文書フィルタリングシステム及び方法は、レイアウトが固定されているが内容が変化している特定のページをユーザがモニターしているこれらのもの状態において理想的である。理由は、これらのユーザがそのページのフォーマットに対してフィルタを同調できるからである。

【0024】本発明の自動再オーサリング及び文書フィルタリングシステム及び方法は、限られた通信帯域幅及び小さな表示を有する装置に、インターネットやイントラネットのような分散形ネットワーク上の任意の文書へのアクセスを提供するために、文書フィルタリングに連結された自動再オーサリング能力を提供する。

【0025】本発明の自動文書再オーサリング及び文書フィルタリングシステム及び方法は、分散形ネットワークから文書に対する要求を捕獲して、元に対する要求された文書ではなくて要求された文書の再オーサリング済バージョンを戻す。

【0026】モバイル及びユービキタス(ubiquitous)コンピュティングのより大きなコンテキスト(文脈)において、本発明の自動文書再オーサリング及び文書フィルタリングシステム及び方法は、ユーザのビューモビリティをプラットフォームに与えるためのキーテクノロジーを提供する。本発明の第1の態様は、文書の望ましいエレメントを抽出するために前記文書をフィルタリングする方法であって、所定のフィルタを提供するステップと、フィルタリングされるべき前記文書を得るステップと、前記得られた文書から解析ツールを生成するステップと、前記提供されたフィルタに基づいて前記解析ツールを介してナビゲートして抽出されるべき前記文書の少なくとも一つのエレメントを識別するステップと、前記少なくとも一つの識別されたエレメントを抽出するステップと、前記少なくとも一つの抽出されたエレメントを組合せて前記文書のフィルタリングされたバージョンを形成するステップと、を備える文書のフィルタリング方法である。第2の態様は、第1の態様において、前記所定のフィルタを提供するステップは、分散形ネットワークのノードから前記所定のフィルタを検索するステップを有する。第3の態様は、第1の態様において、前記所定のフィルタを提供するステップは、前記所定のフィルタに対するユニフォームリソースローケータ(URL)を分散形ネットワークのホストノードへ提供するステップと、前記提供されたURLに対応して、前記分散形ネットワークのノードから前記所定のフィルタを検索するステップと、を備える。第4の態様は、第3の態様において、前記所定のフィルタに対するURLを提供するステップは、前記URLを前記分散形ネットワークに接続されたデータ処理装置から送信するステップを備え

る。第5の態様は、第4の態様において、前記データ処理装置から前記URLを送信するステップは、前記URLを無線通信チャネルを介して送信するステップを備える。第6の態様は、第3の態様において、前記分散形ネットワークの前記ノードから前記所定のフィルタを検索するステップは、前記分散形ネットワークの前記ホストノードから前記所定のフィルタを検索するステップを備える。第7の態様は、第3の態様において、前記分散形ネットワークの前記ノードから前記所定のフィルタを検索するステップは、前記分散形ネットワークのノードから前記所定のフィルタを検索するステップを備える。第8の態様は、第1の態様において、前記フィルタリングされるべき前記文書を得るステップは、分散形ネットワークのノードから前記フィルタリングされるべき文書を検索するステップを備える。第9の態様は、第1の態様において、フィルタリングされるべき前記文書を得るステップは、前記フィルタリングされるべき文書に対するURLを提供するステップと、前記提供されたURLに対応して、分散形ネットワークのノードから前記フィルタリングされるべき文書を検索するステップと、を備える。第10の態様は、第9の態様において、前記フィルタリングされるべき文書に対する前記URLを提供するステップは、前記フィルタリングされるべき文書に対する前記URLは前記所定のフィルタに埋め込むステップを備える。第11の態様は、第10の態様において、前記フィルタリングされるべき文書に対する前記URLを前記所定のフィルタに埋め込むステップは、前記フィルタに、GO URLコマンドを使用するステップを備える。第12の態様は、第1の態様において、前記解析ツリーを介してナビゲートするステップは、前記解析ツリーの現在のコンテキストに更に移動するステップと前記解析ツリーの現在のコンテキストから外に移動するステップと前記解析ツリー内で現在のコンテキストに相対して前記解析ツリーの他の構造へ移動するステップの内の少なくとも一つを含む。第13の態様は、第12の態様において、前記現在のコンテキストは、前記解析ツリーの現在のノードであり、前記解析ツリーの現在のコンテキストに更に移動するステップは、前記解析ツリーの現在のノードから前記現在のノードの少なくとも一つの子ノードへ移動するステップを備える。第14の態様は、第13の態様において、前記解析ツリーの現在のノードから前記現在のノードの少なくとも一つの子ノードへ移動するステップを備える前記解析ツリーの現在のコンテキストに更に移動するステップは、GO「正則表現」コマンドを使用

するステップを備え、前記少なくとも一つの子ノードは前記「正則表現」によって定義されるストリングを含む。第16の態様は、第12の態様において、前記現在のコンテキストは、前記解析ツリーの現在のノードであり、前記解析ツリーの前記現在のコンテキストから外へ移動するステップは、前記解析ツリーの現在のノードから現在のノードの少なくとも一つの親ノードへ移動するステップを備える前記解析ツリーの現在のコンテキストから外へ移動するステップは、GO ENCLCINGコマンドを使用するステップを備える。第18の態様は、第17の態様において、前記GO ENCLCINGコマンドを使用するステップは、GO ENCLCING「構造」コマンドを使用するステップを備え、前記「構造」は、現在のノードへ導く解析ツリーのノードである。第19の態様は、第17の態様において、前記GO ENCLCINGコマンドを使用するステップは、GO ENCLCING「正則表現」コマンドを使用するステップを備え、前記少なくとも一つの親ノードは、前記「正則表現」によって定義されるストリングを含む。第20の態様は、第12の態様において、前記現在のコンテキストは、前記解析ツリーの現在のノードであり、前記解析ツリー内で現在のコンテキストに対して前記解析ツリーの次の構造へ移動するステップは、現在のノードから前記解析ツリーのノードを介して移動するステップを備える。第21の態様は、第20の態様において、現在のノードから前記解析ツリーのノードを介して移動するステップは、GO NEXTコマンド、GO PREVIOUSコマンド、GO FIRSTコマンド、GO LINKEDPAGEコマンド、及びこれらのコマンドの組合せの内の少なくとも一つを使用するステップを備える。第22の態様は、第21の態様において、GO NEXTコマンド、GO PREVIOUSコマンド、GO FIRSTコマンド、GO LINKEDPAGEコマンド、及びこれらのコマンドの組合せの内の少なくとも一つを使用する前記ステップは、GO NEXT「構造」コマンド、GO PREVIOUS「構造」コマンド、GO FIRST「構造」コマンドの少なくとも一つを使用するステップを備え、前記「構造」は、前記解析ツリーのノードである。第23の態様は、第21の態様において、GO NEXTコマンド、GO PREVIOUSコマンド、GO FIRSTコマンド、GO LINKEDPAGEコマンド、及びこれらのコマンドの組合せの内の少なくとも一つを使用する前記ステップは、GO NEXT「正則表現」コマンド、GO PREVIOUS「正則表現」コマンド、GO FIRST「正則表現」コマンドの少なくとも一つを使用するステップを備え、宛先ノードは、前記「正則表現」によって定義されるストリングを含む。第24の態様は、第20の態様において、現在のノードから前記解析ツリーのノードを介して移動する前記ステップは、現在のノードから前記解析ツリーを介してシーケンシャルに移動するステップを

備える。第25の態様は、第12の態様において、前記解析ツリーを介してナビゲートする前記ステップは、前記検索されたフィルタで定義された正規表現とのマッチングに基づいて、現在のコンテキストへの移動、現在のコンテキストから外への移動、及び現在のコンテキストに相対する移動の内の何れかを更に備える。第26の態様は、第12の態様において、前記解析ツリーを介してナビゲートする前記ステップは、前記検索されたフィルタで定義された正規表現との反復マッチングに基づいて、現在のコンテキストへの移動、現在のコンテキストから外への移動、及び現在のコンテキストに相対する移動の内の何れかを更に備える。第27の態様は、第26の態様において、前記正規表現の反復マッチングは、PO REACH[®] 正規表現[®]を使用するステップを備え、宛先ノードは、前記“正規表現”によって定義されるストリングを含む。第28の態様は、第12の態様において、前記解析ツリーを介してナビゲートする前記ステップは、前記検索されたフィルタで定義されたナビゲーションコマンドの反復実行に基づいて、現在のコンテキストへの移動、現在のコンテキストから外への移動、及び現在のコンテキストに相対する移動の内の何れかを更に備える。第29の態様は、第28の態様において、前記検索されたフィルタで定義されたナビゲーションコマンドの反復実行は、FOREACH[®] 構造[®]コマンドを使用するステップを備え、宛先ノードは、前記“構造”によって定義されるストリングを含む。第30の態様は、第1の態様において、前記解析ツリーを介してナビゲートする前記ステップは、前記解析ツリー内の各望ましい位置を構造的に記述することに基づいて、前記解析ツリー内での少なくとも一つの望ましい位置へ移動するステップを備える。第31の態様は、第1の態様において、前記解析ツリーを介してナビゲートする前記ステップは、各望ましい位置のコンテキスト内に含まれる正規表現を記述することに基づいて、前記解析ツリー内での少なくとも一つの望ましい位置へ移動するステップを備える。第32の態様は、第1の態様において、更に、前記少なくとも一つの前記抽出されたエレメントを含む文書を生産するステップと、前記生成された文書から評価値を生産するステップと、前記評価値が少なくとも一つの評価基準を満足するか否かを決定するステップと、前記評価値が前記一つの評価基準を満足する場合、前記文書を出するステップと、前記評価値が前記一つの評価基準を満足しない場合、前記文書を変形済文書へ変形して、前記生成するステップ及び前記決定するステップを繰り返すステップとを備える。第33の態様は、文書から少なくとも一つのエレメントを抽出するために、選択されたフィルタに基づいて、前記文書を自動的にフィルタリングする文書フィルタリングシステムであって、解析ツリー生成メモリと、文書フィルタリングメモリと、抽出されたエレメント出力メモリと、を備える文書フィルタリングシステムである。

第34の態様は、第33の態様において、前記解析ツリー生成メモリは、解析ツリーを生産するために前記文書を解析する。第35の態様は、第34の態様において、前記解析ツリーは、抽象構文ツリーである。第36の態様は、第34の態様において、前記文書フィルタリングシステムは、前記選択されたフィルタのナビゲーションコマンドに基づいて、前記解析ツリー内でナビゲートする。第37の態様は、第34の態様において、前記文書フィルタリングシステムは、前記選択されたフィルタ内レポートコマンドと前記解析ツリーの現在のコンテキストに基づいて、前記文書から文書エレメントを抽出する。第38の態様は、第37の態様において、前記文書フィルタリングシステムは、前記選択されたフィルタのナビゲーションコマンドに基づいて、前記解析ツリー内でナビゲートすることによって、前記解析ツリーの現在のコンテキストを決定する。第39の態様は、第33の態様において、更に、文書サイズ評価メモリと、変形メモリと、を備える。第40の態様は、第39の態様において、前記文書サイズ評価メモリは、前記少なくとも一つの抽出されたエレメントを含むフィルタリングされた文書から前記解析ツリー生成メモリによって生成された解析ツリーを評価して、前記文書が少なくとも一つの評価基準を満足するか否かを決定する。第41の態様は、第40の態様において、前記フィルタリングされた文書が前記少なくとも一つの評価基準を満足する場合、前記フィルタリングされた文書が、デスクトップモニターの表示領域よりも小さい表示領域を有する表示装置へ出力される。第42の態様は、第41の態様において、前記フィルタリングされた文書が前記少なくとも一つの評価基準を満足しない場合、前記変形メモリは、第1の変形を使用して前記解析ツリーを変形して第1の変形済解析ツリーを生産する。第43の態様は、第42の態様において、前記文書サイズ評価メモリは、前記変形メモリによって生成された変形済解析ツリーを評価して前記変形された解析ツリーに対応する変形された文書が少なくとも一つの評価基準を満足するか否かを決定する。

【0027】本発明のこれらの及び他の特徴及び利点が以下の好ましい実施の形態の詳細な説明において述べられると共にそれから明らかとなる。

【0028】

【発明の実施の形態】本発明の文書フィルタリングシステム及び方法の以下の説明において、用語“ウェブページ”、“ウェブ文書”及び“文書”は、インターネット、インターネット、インターネットのワールドワイドウェブ部分やあらゆる他の既知の或いは将来開発される分散形ネットワークのような、分散形ネットワークからの一つのエンティティとして検索されるあらゆるセットの情報を含むことを意図される。この情報は、テキストストリング、画像、テキストストリングと画像の表、他

のウェブページへのリンク、テキストストリング、画像、ウェブページ内の表及びリンクのレイアウトを定義するフォーマット情報を含む。

【0029】多くの可能な自動文書再オーサリング技術があり、それらは、二つのディメンション、即ち、構文技術対意味技術及び変形技術対省略技術、に沿って分類される。構文技術は、文書の構造上で働き、意味技術

は、内容のある程度の理解に依存する。省略技術は、基本的に、ある情報を除去すると共にタッチされないすべての残し、変形技術は、文書の表示又は内容のある態様を変更することを含む。表1は、各カテゴリの例と共にこれらのディメンションを示す。

【表1】

	省略	変形
構文的	セクションのアウトライン化	画像の縮小
意味的	重要でない内容の除去	テキストの要約

異なるタイプの自動再オーサリング技術の例

【0030】自動化された文書再オーサリングシステムによって必要な処理の理解をするために、一般的なウェブページの特性を評価し、手作業による幾つかのウェブページの再オーサリングの処理を介して候補再オーサリング技術を識別するための研究が行われた。

【0031】初めに、“典型的な”ウェブページ、ゼロックスコーポレート (Xerox Corporate) ウェブサイト、の収集が研究の焦点を定めるために選択された。3188のウェブページのこの収集は、最新の専門的にデザインしたウェブサイトを代表する。一般的なページの構造及び内容の理解を得るための助けとして、種々の統計量は、これらのページに関して、ウェブクローラー (web crawler) を使用して収集された。これらの統計量は、全体のウェブに亘って実行された他の大規模な研究と略一致する。

【0032】次に、ゼロックスウェブサイトのサブセットのページが手作業での再オーサリングのために選択された。ゼロックス1995年次報告からの一セットのページが選択され、320x240画素スクリーンを有するシャープザウルス (Sharp Zaurus) PDAに表示するために手作業で変換された。使用された設計方法及び技術の詳細なノート (注記) が維持された。

【0033】この処理の間に学習されたデザインヒューリスティックス (発見) の幾つかは以下の通りである。

(1) 原画像の少なくとも幾つかを維持することは、原文書の見栄え及び感じを維持するために重要である。共通の技術は、第1の画像のみ、又は最初及び最後の画像、即ち、ブックエンド画像、を維持して残りの画像を省略することを含む。

(2) セクションヘッダ、即ち、HTMLのH1-H6タグは、正確にはそれほどしばしば使用されない。それらのヘッダは、使用される場合は、例えば、ボールドのような、特定のフォントサイズ及びスタイルを達成するために、かなりしばしば使用される。このように、セクションヘッダは、大部分の書類のための構造的アウトラインを提供するためには、信頼できない。代わりに、多くのテキストブロックを有する文書は、各テキストブロック

を第1のセンテンス又は各ブロックのフレーズで置き換えること、即ち、第1のセンテンスの削除によって減少される。

(3) 画像に対する最初の処理は、標準の割合で全ての画像のサイズを縮小して、文書がオーサリングされた表示領域の目標設置の表示領域に対する比率によって記述することである。

(4) 意味省略は、ページに表される主概念に付随する情報を表すサイドバーに対して実行される。ゼロックスページの多くがそのようなサイドバーを備え、それらは、単純に減少 (縮小) されたバージョンでは削除された。

(5) また、意味省略は、情報をそのページに貢献させないが、その審美性を改良するのに働く画像に実行される。

(6) ページは、カテゴリ化され、次にそれらのカテゴリに基づいて再オーサリングされる。これらの二つの例は、バナーとリンク表である。バナーは、主に一セットの画像と、少数のナビゲーションリンク (これは一つのみが多い) とを含み、それらのリンクは、審美的外観を構築するように働くが、殆ど又は全く内容を含まない。スペースがプレミアムの場合 (顔面以上である場合)、それらは、通常全体的に省略される。リンク表ページは、主に、他のページのハイパーテキストリンクのセットであり、従って、ほんの僅かな追加の内容を含む。これらのリンクページは、通常、次のブロックのそれらのリンクをリストアップする、よりコンパクトなフォームに再フォーマットされる。

(7) 大きな表示上では価値無い白スペースが小さな装置では貴重である。幾つかの技術は、ページ中の白スペースの量を減少するために発見された。パラグラフのシーケンス、即ち、HTML “P” タグやブレイク (区切り)、即ち、HTML “BR” タグ、が一つのそのようなパラグラフやブレイクに縮小される。リスト、即ち、HTML “UL”、“OL” 及び/又は “DL” タグ、はそれらのインデント (字下げ) 及びバレット (黒丸) を有する価値のある水平方向スペースを取る。これらのリストは、クー

パー (Cooper) らが述べたように、連続する項目 (アイテム) 間のブレイクを備える簡単なテキストブロックに再フォーマットされる。

【0034】結論として、文書再オーサリングを実行するためには、二つの事が必要である。即ち、一セットの再オーサリング技術、即ち、一セットのページ変形と、ページ変形を適用するための戦略とである。手作業による再オーサリングの研究で使用される技術の内の成文化のために最も大切な技術は、セクションアウトライン化、第1センテンス省略及び画像省略を含む構文省略技術と、画像サイズ縮小及びフォントサイズ縮小を含む構文変形技術であった。この研究の間に学んだデザイン戦略は、変形技術のリンク付け、即ち、種々のトリアル、及び各変形又は変形の組合せが適用されるべき条件のセットを含んでいる。

【0035】上述された研究の結果に続いて、本発明の文書再オーサリングシステム及び方法に対して二つの大きな要素がある。即ち、種々の方法で文書を変形する個々の再オーサリング技術の取集と、所与の文書/表示サイズ対のために技術の最良の組合せを選択することによってデザイン戦略を実施する自動化文書再オーサリングシステム及び方法である。

【0036】セクションヘッダアウトライン化 (Section Header Outlining) 変形は、技術論文及びレポートのような、構造的な文書のために必要な表示サイズを減少するための優れた方法を提供する。このアウトライン化処理は図1に示されている。

【0037】図1に示されるように、文書100は、セクションページ110のリストに変換され、各セクションは、ページ111の中に省略される。即ち、文書100の各セクション102の内容106は、文書100から省略され、各セクションヘッダ104がハイパーテキストに変換される。何らかのセクションに対するハイパーテキストリンクが選択されると、省略された内容の対応するページ (セクション112) がブラウザにロードされる。複数のセクションレベル (セクション、サブセクション、サブセクションのサブセクション、等) に直面すると、省略を実行するための二つのアプローチがある。第1のアプローチは、完全なアウトライン化であり、それは、セクションヘッダのみが維持され、その結果は、本の目次のように見える。第2のアプローチは、レベル (to-level) アウトライン化である。このレベルへのアウトライン化において、セクション階層のカットオフレベルが決定され、より低いレベルのセクションヘッダを含む、そのレベルより下のすべての内容が省略されるが、それより上の全ての内容が維持される。

【0038】大部分のページがテキストブロックを有するので、セクションヘッダが存在しない場合でさえ、第1センテンス省略 (First Sentence Ellision) 変形は、必要なスクリーン領域を減少する優れた方法であり

得る。この技術において、各テキストブロックは、その第1のセンテンス、或いは自然のブレイク点までのその第1のフレーズで置き換えられる。この第1のセンテンス又はフレーズは、またものとテキストブロックに対するハイパーテキストリンクとされる。

【0039】インデックス付けされたセグメント (Indexed Segment) 変形は、初めに、順序付けされた或いは順序付けされていないリスト、パラグラフや表のシーケンスのような、論理的に区分され得るページ要素を見つけることを試みる。この変形は、入力ページを受け取り、項目のある数を各々のサブページに割り当てることによって、その内容をサブページにセグメント化し、インデックスページを構築してそれをサブページの集まりに対して準備する。次に、インデックス付けされたセグメント変形は、各ページがクライアントの表示サイズに対して“完全”であるまで、出力ページにこれらの要素を順序通りに充填することを開始する。単一の論理要素が単一の出力ページに当てはまらない場合、次に、インデックス付けされたセグメント変形は、パラグラフ又はセンテンス境界上のテキストブロックを区分する第2の区分化を実行する。

【0040】インデックス付けされたセグメント変形において、可能な限り多くのスタイル情報は、その祖先区分のHTMLタグの全ての中に埋め込まれた各エレメント (要素) を出力することによって、出力エレメントに対して保持される。次に、インデックス付けされたセグメント変形は、出力されるべき各エレメントからセクションヘッダ又は第1のセンテンスをコピーし、コピーされた部分をインデックスページに連結し、各コピーされた部分から適切なサブページのハイパーテキストリンクを生成することによってインデックスページを構成する。インデックスページ自体が、区分化される必要がある場合があることを理解すべきである。インデックス付けされたセグメント変形において、更に、連続するサブページ間の“Next”及び“Previous”ナビゲーションリンクがナビゲーションの便宜のために追加される。

【0041】表 (Table) 変形は、ページ上の表、即ち、矩形格子状に配置された情報の表示、がクライアントに直接送られない時に、認められる。これらの場合、表変形は、上から下へ、左から右への順序を使用し、表のセル当り一つのサブページを生成する。表に入力された表は、同様に処理される。表コラムが市販のHTMLウェブページで一般に行われている“ナビゲーションサイドバー”として使用されている時を決定するために、表変形はヒューリスティクスを使用する。この場合、表変形は、それらのセルが3人の僅かな内容を支持することを意図するので、これらのセルをサブページのリストの最後に移動する。

【0042】図2は、入力表を示し、表を表セルよりも太い境界でマークする。図2に示される表120にお

いて、セル 1 2 2 は、サイドバーとして識別され、それは、セル 1 2 8 の後に配される。他のセルの全ては、それらの自然の順序で配される。サブセル 1 2 5 及び 1 2 6 のような、セル 1 2 4 の六つの部分は、それらが白スペースのみしか含まない場合、各々サブセル 1 2 3 及び 1 2 7 を含むサブページ間とのそれら自体のサブページに配される。

【 0 0 4 3 】 この例から理解されるように、入れ子テーブル及びサイドバーは、表の処理を複雑にする。これは、そのサイドバーが内側の表の部分である場合、特に当てはまる。この状態で、サイドバーは、何れかの取り囲む表の最後にはではなくて、内側の表の最後へ移動されるべきである。本発明の文書再オーサリングシステム及び方法の例示の実施の形態において、表によってセルをグループ化するのではなくて、サイドバーが一度に一つの表に移動されて、次に全ての表セルが一度に処理される。

【 0 0 4 4 】 画像が存在することは、自動文書再オーサリングにとって最も困難な問題の一つである。その理由は、所与の画像を維持、縮小、又は削除されるべきか否かの決定が、そのページの画像の内容及び役割の理解に基づくべきであるという点である。しかしながら、画像縮小 (Image Reduction) 変形及び画像省略 (Image Elision) 変形は、ユーザが原画像を検索できるメカニズムがユーザに提供される限りにおいて、内容の理解無しに、適用できる。本発明のシステム及び方法の例示の実施の形態において、画像縮小変形は、2 5 %、5 0 % 及び 7 5 % のような一セットの予め定義されたスケールリングファクタの一つによって一ページ中の全ての画像を縮小し、縮小された画像を原画像にリンクするハイパーテキストリンクにする。

【 0 0 4 5 】 画像縮小変形に加えて、三つの意味省略 (Syntactic Elision) 変形もまた画像のために開発された。全て省略 (Elide All) 変形、第 1 画像のみ (First Image Only) 変形、及びブックエンド (Bookends) 変形である。全て省略変形において、その文書から全ての画像が省略される。第 1 画像のみ変形において、第 1 の画像を除いて全てが文書から省略される。ブックエンド変形において、第 1 及び最後の画像以外の全てが文書から省略される。省略された画像の各々は、利用可能な場合、それらの HTML ALT テキストで置き換えられる。ALT テキスト又は各省略された画像に対する標準のアイコンもまたその原画像に対するハイパーテキストリンクにされる。

【 0 0 4 6 】 本発明の文書再オーサリングシステム及び方法の一例示の実施の形態において、スクリーンスペースがあまりに制限されている場合又はクライアント装置が画像を表示できない場合、これらの画像が文書から除去される。しかしながら、除去された画像は、クライアントサイド画像マップを介するハイパーテキストリンク

のためのアンカーとして使用され得る。そのような画像が除去される場合、HTML 文書によって表されるウェブページがナビゲーション不能にレンダリングされ得ることが理解されるべきである。これを可能とするために、本発明の文書再オーサリングシステム及び方法の一例示の実施の形態において、そのような画像からハイパーテキストリンクを抽出してそれらをリンクアンカーのテキストリストにフォーマット化する変形が使用される。このテキストリストのラベルは、存在する場合、画像マップの HTML ALT タグから成り、そのリンクの URL (一様資源ロケータ) の部分から抽出される。この変形は、画像を除去する場合、ナビゲーションのために画像に取付られるリンクを保存する。

【 0 0 4 7 】 所与のクライアント表示のためにどのような組合せの変形を所与のページへ適用するかを決定するための全体の処理は、初めに人間の審美能力のあるフォーム (形) を必要とするように思える。しかしながら、本発明の自動文書再オーサリングシステム及び方法は、手作業による再オーサリング実行で使用するヒューリスティクス (発見) の多くを捕獲し、所与の表示のための良く見えるページを生成する非常に優れた仕事をを行う。

【 0 0 4 8 】 個々のページ変形がそれらの好ましさによって順序付けされる。変形のどの組合せが所与の文書に適用されるべきかを決定するために、本発明の文書再オーサリングシステム及び方法は、変形のための予備条件及び変形の組合せを記述する多くのヒューリスティクスを使用して、文書変形スペースのデプスファースト (depth-first) 検索を実行する。デプスファースト検索は、最も望ましい変形の組合せを使用することによって、文書の "十分に良い" バージョンが検出されることを保証する。より望ましい変形が適用できない場合や文書を十分に縮小できない場合のみ、あまり好ましくない変形が使用される。

【 0 0 4 9 】 本発明の文書再オーサリングシステム及び方法は、一番優れた方法で文書変形スペースを検索する。この検索スペースの各状態は、文書の一つのバージョンを表し、最初の状態は、"オーサリングされたまま" の文書を表す。各状態は、その状態でタグバージョンの質を表すメリットの尺度を表す数値でタグ付けされる。各状態に対するメリットの尺度、即ち、評価関数又は評価値は、文書がその状態に存在するままに、全体の文書を表示することを要求されるスクリーン領域の大きな推定である。一つの状態は、その状態に存在するように単一の変形技術を再オーサリング済文書に適用することによって、後続の状態に拡張される。

【 0 0 5 0 】 検索処理の全てのステップにおいて、文書の最も見込みのある状態、即ち、最も小さい現在の表示領域要求を有する状態、が選択され、可能な場合は、変形は、文書の現在の状態からその文書のより見込みのあ

る状態へその文書を変形するために適用される。十分に良い”文書バージョンを含む状態が生成されるやいなや、検索が中断でき、その文書のそのバージョンがレンダリングのためにクライアント装置へ戻される。或いは、検索は、原ページの全ての内容が含まれるカーセットの十分に優れたサブページに表されるまで、続けられる。検索が行われ尽くされ、十分に優れた文書バージョンが発見されない場合、この検索の間に発見された最良の文書が、レンダリングのために、クライアント装置へ戻される。最良の文書によって満足されない厳しいサイズ制約がある場合、文書をバラグラフの中でブレイクアップするより分解的な変形が適用される。

【0051】図3は、文書200に異なる変形を適用して異なる結果となる再オーサリング済サブページ210、220及び230を生成する方法を示す。ユーザによって本発明のシステム及び方法へ供給される情報に依存して、サブページ210、220及び230の一つが”最良の”再オーサリング済ページとして選択される。次に、例えば、第1のサブページから除去された内容のために十分に優れたサブページを生成するために更なる再オーサリングが必要場合、或いは最良のサブページがまだ”十分に良い”でない場合、選択された最良の再オーサリング済サブページ210、220又は230から得られたサブページに或いは選択された最良の再オーサリング済サブページ210、220又は230を再オーサリングするために、追加の変形が適用され得る。

【0052】ヒューリスティック情報が本発明の文書再オーサリングシステム及び方法によって次の幾つかの場合に使用される。それらの場合は、種々の変形技術が所与の状態に適用される順序、各変形技術に対する予備条件、及び文書のバージョン又はサブページが”十分に良い”である時の決定を含む。一般に、その文書を僅かに変化する変形は、より大きな変化を行う変形よりも好ましい。例えば、25%だけ画像を減少することは、それらの画像を75%縮小することよりも好ましい。

【0053】各変形技術に対する予備条件は、その変形

```

Digestor(initial_page)
to_be_reauthored={initial_page}
to_deliver={}
while(to_be_reauthored!={})
next_page=pop(to_be_reauthored)
best_version_state=reauthor(next_page)
to_deliver.append(best_version_state.page)
to_be_reauthored.append(best_version_state.sub_pages)
return to_deliver

```

【0056】全ての再オーサリング済サブページは変形された解析ツリーとしてキャッシュされる。ユーザが変形された文書をナビゲートしてサブページを要求すると、対応する解析ツリーがレンダリングされてクライ

ア組み合わされ得るその他の変形を指定する。例えば、フルアウトライン化及び第1のセンテンス省略の両方を同じ文書に適用することは、意味がない。また、予備条件は、その技術が適用されている文書の内容及び構造についての必要条件を指定する。例えば、フルアウトライン化(Full Outlining)変形は、再オーサリング待ちその文書又はサブページに少なくとも三つのセクションヘッドがある場合にのみ適用されるべきである。十分に良い”のための現在の条件は、非常に単純化されている。即ち、検索は、文書又はサブページによって必要とされる領域がクライアント表示の所定のマルチプル(複数)のスクリーン領域である場合に、停止される。一般に、この所定のマルチプルはより大きく、一実施の形態では、2.5である。このより高いマルチプルは、ユーザが一つの方向へほんの僅かでも表示をスクロールしなくないことを仮定するに過ぎない。

【0054】図2の示されるように、変形が文書に適用されると、文書が複数のより小さな”サブページ”に分割され得る。しかしながら、これらのサブページの各々は、まだ大き過ぎてダウンロードしてクライアント表示上に表示できないかもしれない。この問題を解決するために、本発明の文書再オーサリングシステム及び方法は、結果としての文書バージョンを表示する状態に取付けられる各シーケンスの変形によって生成されるサブページのリストを保つ。実際にはクライアントへ送出される第1サブページの十分に優れたバージョンのみである文書の十分に優れたバージョンが選択されると、そのバージョンに対する生成されたサブページのリストが再オーサリング待ちページの全体のリストへ追加される。次に、本発明の文書再オーサリングシステム及び方法は、結果としてのサブページの全てがクライアントへ送出されるまで、これらの再オーサリング待ちページの各々を再オーサリングする。この手順は、以下の疑似コードで示されており、そこで”再オーサリング”は、単一の入力ページに対する上述の最良再オーサリング処理と呼ばれる。

【0055】

ントへ送られる。

【0057】本発明の文書再オーサリングシステム及び方法は、最初に文書を解析し、その文書の解析ツリー又は抽象構文ツリー(AST)表現を構成することによって

文書を再オーサリングする。次に、本発明の文書再オーサリングシステム及び方法は、一連の変形を解析ツールに適用する。次に、本発明の文書再オーサリングシステム及び方法は、各得られる変形された解析ツールを文書表現にマッピングし返す。それは、原文書の入力フォーマットとは異なる文書フォーマット内にあり得る。

【0058】文書バージョンベース内の一つの状態ノードを取りその変形がその状態に適用されるべき場合に真に戻る状態関数と、その変形が実際に一つの状態に適用されて、新たな文書バージョン、新たな品質の尺度、及び得られるサブページを含む新たな一つの状態を生成する時に呼び出されるアクション関数と、を含む、標準の手順を使用して、文書変形が実行される。変形の三つのタイプが定義され得る。一つは、プランニング処理が開始する前に、常にページ上に実行される変形であり、二つ目は、最良のプランニング処理で実行される変形であり、三つ目は、最終の抽象構文ツリーからHTMLのような表面フォームに変形し返される前に、常にページ上で実行される変形である。

【0059】変形は、文書の新たなバージョンを生成するために、変形が適用される状態において、解析ツールを操作する。それらの操作は、1996年5月、フランスのパリにおいて、第5回国際ワールドワイドウェブ会議でのS. Bonhomme等による"Interactively Restructuring HTML Documents"に記述されたものと同様である。解析ツリーの複数の部分が省略又は変形される時は常に、全ての影響された解析ツリーのサブツリーのノード識別子を参照するために、HTMLハイパーテキストリンクは解析ツールに追加され、ユーザが再オーサリングの間に変更された文書の原部分を要求することが可能である。

【0060】また、本発明の文書再オーサリングシステム及び方法は、全ての変形が交換可能であると仮定すると、二つの同じ状態が構成されないように、変形のどの組合せがすでに試みられたかをトラック（追跡）を維持する。

【0061】上述のように、本発明の一例示の文書再オーサリングシステム及び方法は、HTTPプロキシサーバとして実施された。HTTPプロキシサーバは、HTML文書に対する要求を受信し、指定されたHTTPサーバから文書を検索し、HTML文書を解析し、検索されたHTML文書から解析ツール即ち抽象構文ツリーを構成し、独自の識別子で解析ツリーノードの各々をラベル付けし、そしてあらゆる埋め込まれた画像を検索し、検索しなば、検索された画像のサイズが決定され得る。これが完了すると、本発明の文書再オーサリングシステム及び方法は、原検索された文書に対する解析ツールを含む状態で初期化される。各再オーサリングサイクルの間、本発明の文書再オーサリングシステム及び方法は、可能な限り最良の文書バージョンを有する状態を選択し、次に、最良に適用

できる変形技術を選択し、その選択された変形を適用し、新たな状態及び新たな文書バージョンが生成される。冗長な状態が構成されないように、変形の組み込みが絶えず交換可能であり、幾つかのチェックが本発明の再オーサリングシステム及び方法によって使用されることが行われ得る。

【0062】本発明の文書再オーサリングシステム及び方法の一例示の実施の形態において、次の15の変形技術が実行された。フルアウトライン (FullOutline)、アウトラインツリーH1 (OutlineH1)、アウトラインツリーH2 (OutlineH2)、アウトラインツリーH3 (OutlineH3)、アウトラインツリーH4 (OutlineH4)、アウトラインツリーH5 (OutlineH5)、アウトラインツリーH6 (OutlineH6)、ファーストセンテンス省略 (FirstSentenceElision)、縮小画像25% (ReduceImages25%)、縮小画像50% (ReduceImages50%)、縮小画像75% (ReduceImages75%)、省略全画像 (ElideAllImages)、第1画像のみ (FirstImageOnly)、バックエンド画像 (BackendImages) 及び縮小フォントサイズ (ReduceFontSize) である。

【0063】本発明の文書再オーサリングシステム及び方法の一例示の実施の形態は、ジャバ (Java) プログラミング言語で実行された。真のプロキシサーバとして機能することに加えて、また、このHTTPプロキシサーバシステムは、HTTPプロキシサーバ自体によって生成された文書を有する幾つかのURLに対する要求に応答できる。これは、HTTPプロキシサーバ及び文書再オーサリングシステムと方法へのフォームベースの制御をユーザに提供するために使用される。文書再オーサリングシステムのこの例示の実施の形態は、シマンテックジャバ JIT (Symantec's Java JIT) コンパイラを使用して、200メガヘルツペンティアム (200MhzPentium) 上で2秒未満で非常に複雑なページでさえ処理できる。

【0064】本発明の文書再オーサリングシステム及び方法のユーザが行わなければならない第1のことは、使用される装置の表示のサイズを指定し、使用されるデフォルトブラウザのフォントのサイズを指示する。この情報は、テキストブロックのスクリーン領域要求を予測するために、必要とされる。これを行うために、ユーザは、HTTPプロキシサーバから指定の制御URL（一様資源ロケータ）を要求して、図4に示されるフォーム300を送出する。

【0065】ユーザが文書再オーサリングシステムを構成すると、ユーザは、ワールドワイドウェブ (WWW) のような分散形ネットワークから文書の検索を開始できる。図5に示される原ページ400と再オーサリング済ページ410は、本発明の文書再オーサリングシステム及び方法の再オーサリング能力を示す。この例において、本発明の文書再オーサリングシステム及び方法のこの例示の実施の形態は、原画像400から表示されたベ

ージ410をレンダリングするために、第1のセンテンス省略と組み合わせられて25%画像縮小を使用することを選択する。次に、再オーサリング済ページ410は、ブラウザウィンドウ420上に表示される。本発明の文書再オーサリングシステム及び方法のこの例示の実施の形態において、ページの検索に直ぐ続いて、ユーザは、HTTPプロキシサーバーから他の制御側を要求することによって、どの変形が適用されたかを決定するために、再オーサリングされるセッションのトレースを要求することが出来る。

【0066】図6は、本発明の自動文書再オーサリングシステムと方法及び／又は自動文書フィルタリングシステムと方法が実施され得る環境500の例示の実施の形態を示す。図6に示されるように、環境500は、デスクトップ又はラップトップコンピュータのモニターの表示に比較して非常に限られた(小さい)表示領域を備える表示を有する限られた表示領域装置510を含む。図6に示されるように、更に、環境500は、送受信装置通信システム550、分散形ネットワークのホストノード570、及びその分散形ネットワークの残りの部分590を含む。

【0067】この環境500において、限られた表示領域装置510は、通常、無線通信チャネル530によって送受信装置通信システム550に接続されるパーソナルデジタルアシスタンス(PDA)、セルラ電話等である。従って、図6に示されるように、限られた表示領域装置510は、通常アンテナ520を含み、送受信装置通信システム550は、通常、対応するアンテナ540を含む。限られた表示領域装置510は、通常アンテナ520と540との間に送信される無線周波数信号を使用して、無線通信チャネル530を介して送受信装置通信システム550と通信する。

【0068】送受信装置通信システム550は、無線通信チャネル530を介して限られた表示領域装置510から受信されたアナログ又はデジタル信号を分散形ネットワークのホストノード570によって使用可能なフォーマットに変換する。次に、送受信装置通信システム550は、無線通信チャネル530を介して受信された信号を通信リンク560を介して分散形ネットワークのホストノード570へ送信する。通信リンク560は、送受信装置通信システム550と送受信装置通信システム550との間の適切な信号の送信が可能な既知の或いは将来開発され得る通信構造であり得る。送受信装置通信システム550と通信リンク560の正確な構造は、これらの構成要素がどのように実施されるかに依存する設計の選択事項に過ぎないが、そのような設計の選択は、当業者に容易に明らかとなると共に予測され得るので、これらの構成要素の更なる詳細な記述は省略される。

【0069】また、限られた表示領域装置510は、通信リンク522のような無線通信チャネル530以外に

よって分散形ネットワークのホストノード570へ接続され得る。即ち、通信リンク522は、ローカルエリアネットワーク、ワイドエリアネットワーク、公共交換電話ネットワークを介するモデム接続とケーブルテレビシステム等のあらゆる他の既知の通信構造であり得る。例えば、限られた表示領域装置510のユーザは、無線通信チャネル530ではなくて、限られた表示領域装置510をモデムを使用して、公共交換電話ネットワークへ直接接続できる。次に、ユーザは、分散形ネットワークのホストノード570へ直接ダイヤルする。

【0070】分散形ネットワークのホストノード570が最終的に限られた表示領域装置510にどのように接続されようと、分散形ネットワークのホストノード570が限られた表示領域装置510への文書の送信要求を受信すると、分散形ネットワークのホストノード570は、最初に、それに対する要求された文書が分散形ネットワークのホストノード570上に局所的に検出されるか否かを決定する。要求された文書が局所的に検出されない場合、分散形ネットワークのホストノード570は、その文書を要求するために、通信構造580を介して分散形ネットワークの残りの部分590と通信する。その文書を格納する分散形ネットワークの残りの部分590の特定のノードは、最終的に、ホストノード570から通信構造580を介してそれに対する要求を受信し、要求された文書を通信構造580を介してホストノード570へ戻す。通信構造580は、分散形ネットワークの広く検出されるノードと共にリンクするためのあらゆる既知の或いは将来開発されるであろう通信構造及びプロトコルシステムであり得る。

【0071】分散形ネットワークのホストノード570が、要求された文書を受信すると、分散形ネットワークのホストノード570上で実行するHTTPプロキシサーバーは、限られた表示領域装置510についての予め提供された情報に基づいて、要求された文書を再オーサリングする。次に、最初に再オーサリング済ページは、ホストノード570によって、無線通信リンク530又は通信リンク522を介して限られた表示領域装置510に送信される。ユーザが送出されたページをレビューする時、ユーザは、再オーサリング済ページから除去された追加の情報を見る事を要求するを決定してもよい。この場合、ユーザは、望ましい再オーサリング済ページを得るために、無線通信リンク530又は通信リンク522を介して分散形ネットワークのホストノード570に要求を送信してもよい。これに対する要求に応答して、ホストノード570は、無線通信リンク530又は通信リンク522を介して原文書の更なる再オーサリング済サブページを限られた表示領域装置510に送信する。

【0072】図7は、この情報の流れのより詳細を示す。図7に示されるように、限られた表示領域装置510のユーザが、分散形ネットワーク上に存在する特定の

文書をレビューすることを望む場合、ユーザは、その特定の文書に対する要求を限られた表示領域装置510からその分散形ネットワークのホストノード570に存在するHTTPプロキシサーバ571へ送る。次に、HTTPプロキシサーバ571は、その特定の文書に対する要求をそに対する要求されたページを格納する分散形ネットワーク上の特定の遠隔ノード591へ送信する。この特定の遠隔ノード591は、要求された原文書をそのHTTPプロキシサーバ571に存在する文書再オーサリングシステム600へ戻す。文書再オーサリングシステム600は、各々ができるだけつり表示可能な、限られた表示領域装置510上に表示され得る複数のサブ文書に原文書を再オーサリングする。次に、文書再オーサリングシステム600は、ページに対して最初に再オーサリング済ものを限られた表示領域装置510に送出すると共に、その他の再オーサリング済ページは、文書再オーサリングシステムの再オーサリング済サブページのキャッシュ636に格納される。従って、限られた表示領域装置510のユーザが、再オーサリング済ページのキャッシュ636に格納された再オーサリング済サブページの一つに存在する情報を見ることを望む場合、ユーザは、限られた表示領域装置510にそのサブページに対する要求を送信させる。要求されキャッシュに格納されたサブページは、再オーサリング済サブページキャッシュ636から限られた表示領域装置510へ送出される。

【0073】独立の構成要素として、HTTPプロキシサーバ571、文書再オーサリングシステム600及び再オーサリング済サブページキャッシュ636は、図7に示されると共に、一般に、これらの構成要素は、単一のソフトウェアアプリケーションの異なるモジュールのような、単一のエンティティの異なる部分として実施される。

【0074】図8は、文書再オーサリングシステム600の一実施の形態のより詳細なアウトラインを示す基本ブロック図である。図8に示されるように、文書再オーサリングシステム600は、コントローラ610、入力インターフェース620、メモリ630、抽象構文ツリー生成メモリ640、文書サイズ評価メモリ650、変形メモリ660及びツリーからの文書再マップ (tree-to-document remap) メモリ670を含み、各々は、データ/制御バス680によって相互接続される。図6に関して、上述された通信リンク522、560及び580は、バス入力出力インターフェース620に接続される。

【0075】メモリ630は、原ページメモリ部分631、表示装置サイズメモリ部分632、抽象構文ツリーメモリ部分633、検索スペース部分634、変形メモリ635、図7に関して上述された再オーサリング済ページキャッシュ636、及び再オーサリング待ちサブペ

ージのリスト637を含む、多数の機能的に分離された部分を含む。原ページメモリ部分631は、限られた表示領域装置510によって要求されるページを格納する分散形ネットワークの遠隔ノード591から戻される戻された原文書を格納する。

【0076】表示装置サイズメモリ632は、文書再オーサリングシステム600によって使用される限られた表示領域装置510についての種々のパラメータを得て特定の限られた表示領域装置510のためにページを再オーサリングするために、文書再オーサリングシステム600によって使用される多数のフォーム文書を格納する。また、表示装置サイズメモリ632は、少なくとも一つの限られた表示領域装置510のための特定のサイズパラメータを格納する。限られた表示領域装置510についての種々のパラメータに対する文書再オーサリングシステム600を実施する多くの異なる可能な方法があることを理解すべきである。一つの例示の実施の形態において、文書再オーサリングシステム600は、その限られた表示領域装置510が文書再オーサリングシステム600へ連続して接続されたままである限りにおいて、特定の限られた表示領域装置510のための種々のパラメータを格納できる。この場合、特定の限られた表示領域装置510が文書再オーサリングシステム600に再接続される度に、文書再オーサリングシステム600は、表示領域装置510についての種々のパラメータを得るために使用される種々のフォームを送り、ユーザは、文書再オーサリングシステム600が最初にアクセスされる毎に、これらの種々のパラメータを再送出することを要求する。

【0077】これが表示装置サイズメモリ632に対する要求されたサイズを縮小し、特定の限られた表示領域装置510を識別するためのあらゆるシステムを要求しない場合、このシステムは、限られた表示領域装置510のユーザにより大きな負担を課すか、或いは限られた表示領域装置510から文書再オーサリングシステム600への情報の供給を自動化するための処理を要求する。この自動化は、例えば、限られた表示領域装置510からの情報を要求する文書再オーサリングシステム600によって提供され得る。情報がすでに文書再オーサリングシステム600での前のセッションの間にユーザによって入力されており、その情報がその時に限られた表示領域装置510に格納された場合、ユーザは、その情報の文書再オーサリングシステム600への再供給に能動的に関与する必要はないであろう。

【0078】或いは、情報は、文書再オーサリングシステム600でセッションをはじめた時、ユーザが限られた表示領域装置510から供給され得る識別コードと共に表示装置サイズメモリ632に格納され得る。識別コードを文書再オーサリングシステム600によって、ユーザは、文書再オーサリングシステム600がアクセス

される毎に、限られた表示領域装置 510 についての種々のパラメータの全てを再供給することは要求されない。

【0079】あらゆる場合において、原ページメモリ 631 に格納された原ページを再オーサリングする時に、文書再オーサリングシステム 600 は、上述のように、限られた表示領域装置 510 についての種々のパラメータを使用して、それによって各再オーサリング済ページが、出来る限り近接して、限られた表示領域装置 510 の小さな表示領域に当てはまる。

【0080】抽象構文ツリーメモリ部分 633 は、抽象構文ツリー生成メモリ 640 によって原ページメモリ 631 に格納された原文書から生成された抽象構文ツリーを格納する。変形メモリ部分 635 は、上述の種々の変形、及び各変形が適用され得る条件とどの変形が種々の他の変形と共に使用できないかについての条件を格納する。また、変形メモリ 635 は、任意の特定の変形を特定の再オーサリング済ページに適用することの望ましさの指示を格納する。即ち、上述のように、種々の変形は、少くも画像を縮小するような、より大きく限られた変形を、大きく画像を縮小又はその画像を完全に除去するような、より強力な変形又は上へ適用することを強調する一般的な命令を有する。

【0081】再オーサリング済ページキャッシュ 636 は、文書サイズ評価メモリが、表示装置サイズメモリ 632 に格納された限られた表示領域装置 510 についての種々のパラメータに基づいて、特定の再オーサリング済ページ又はサブページに対する抽象構文ツリーが十分に良いことを指示すると、各再オーサリング済ページ又はサブページに対応する抽象構文ツリーを格納する。再オーサリング待ちサブページのリスト 637 は、原文書又はより早いサブページを変形することによって生成されたこれらのサブページに対する抽象構文ツリーを格納する。これらのサブページは、一般的に、あらゆる縮小サイズの画像又はあらゆる省略された画像の原画像、及び内容が省略されたテキストセグメントの完全なテキストを含む。

【0082】最後に、検索スペースメモリ 634 は、それが、現在操作されている検索スペースの特定の状態に基づいて、原ページメモリ 631 に格納されている原文書又は再オーサリング待ちサブページのリスト 637 に格納された種々のサブページに変形メモリ 635 に格納された種々の変形を適用する時、変形メモリ 660 によって生成された多くの状態を格納する。

【0083】特に、検索スペース 634 における各状態 *i* は、評価値部分、変形された抽象構文ツリー部分及びサブページリスト部分を含む。評価値部分は、文書サイズ評価メモリ 650 によって生成されたその状態 *i* に対する再オーサリング済ページ又はサブページに対して生成された評価値を格納する。変形された抽象構文ツリー

一部分は、変形メモリ 635 中の変形の一つをその状態 *i* の親状態へ適用することによって変形メモリ 660 によって生成された状態 *i* に対する変形された抽象構文ツリーを格納する。サブページリスト部分は、変形メモリ 660 がその状態 *i* を生成するために使用される特定の変形を適用すると、その状態 *i* に対応するページから除去されたあらゆる原内容を格納するために生成されたサブページのリストを格納する。

【0084】状態 0 が原ページメモリ 631 に格納された原文書に対応することが理解されるべきである。特に、状態 0 の評価値部分は、あらゆる再オーサリング前に原文書に対して生成された評価値に対応する。この状態 0 において、変形された抽象構文ツリー部分は、原文書に対する抽象構文ツリー生成メモリによって生成された変形前抽象構文ツリーを格納する。最後に、状態 0 の前に、原文書が原情報の全々を含む、従ってサブページが要求されない場合、サブページリストは空である。

【0085】図 9 は、検索スペースメモリ部分 634 に格納された種々の状態を示す。特に、図 9 は、セクションヘッダ、テキストパラグラフ及び画像を含む文書を示す。図 9 に示されるように、初期状態、即ち状態 0 において、原文書が変形されなかった。また、この初期状態は、原文書に対して生成された、原レーティング、即ち、評価値を示す。また、図 9 は、“全ての画像省略 (elide all images)” 変形は状態 0 の文書を適用することによって状態 0 から生成された状態 1 を示す。状態 1 に示されるように、状態 1 の再オーサリング済サブページは、セクションヘッダ及びテキストを含むが、画像を含まない。むしろ画像の位置に、状態 1 の再オーサリング済サブページは、状態 1 の再オーサリング済サブページから省略された画像を格納するサブページへ状態 1 の再オーサリング済ページをリンクする“IMG”とラベル付けされたリンクを含む。また、状態 1 は、この再オーサリング済文書の評価値を示す。図 9 に示されるように、再オーサリング済ページのサイズ要求が原再オーサリングされていないページのサイズ要求の 4 分の 1 である。

【0086】また、図 9 は、二つの追加の状態、即ち、状態 2 と状態 3 は、状態 0 の文書へ他の変形を適用することによって生成されたことを示す。最後に、図 9 は、状態 1 の再オーサリング済文書を又は状態 1 のサブページに追加の変形を適用することによって三つの状態、即ち、状態 4、状態 5 及び状態 6 を示す。例えば、画像を含むサブページが限られた表示領域装置 510 上に表示するのにはあまり大きく過ぎる場合、“25%縮小”、“50%縮小”、又は“75%縮小”変形を画像に適用して限られた表示領域装置 510 上に表示するのに十分に優れた再オーサリング済画像を得ることによって、中間サブページが生成される。

【0087】さて、動作において、図 8 の文書再オーサ

リングシステム600は、通信リンク580を介して戻された原文書を受信する。受信された又は一般の文書は入出力インターフェース620を介して入力され、コントローラ610の制御下で原ページメモリ631に格納される。次に、コントローラ610の制御下で、抽象構文ツリー生成メモリ640は、原ページメモリ部分631から原文書を入力し、その原文書から抽象構文ツリーを生成する。次に、抽象構文ツリー生成メモリ640によって生成された抽象構文ツリーは、コントローラ610の制御下で、メモリ630の抽象構文ツリーメモリ部分633に格納される。

【0088】次に、コントローラ610の制御下で、文書サイズ評価メモリ650は、原ページメモリ631に格納された原文書に対応する抽象構文ツリーと、再オーサリング済文書が戻されるべき特定の限られた表示領域装置510についての表示サイズメモリ632からの種々のパラメータを入力する。次に、文書サイズ評価メモリ650は、評価値を生成し、検索スペースメモリ部分634の状態0にその評価値を格納する。また、文書サイズ評価メモリ650は、状態0の文書が通信リンク522又は560を介して限られた表示領域装置510に出力するのに十分に良いものであるか否かのコントローラ610の指示を出力する。原文書がすでに十分に良いである場合、その原文書は、更なる変形を受けることなく、直ちに戻される。

【0089】次に、コントローラ610の制御下で、変形メモリ660は、その状態に対する抽象構文ツリーによって表されるように、状態0の文書を入力し、変形メモリ635に格納された変形の一つを入力状態の抽象構文ツリーに適用する。特に、最初に変形メモリ660は、現在の状態iに対して、選択された変形がその文書の現在の状態iへ適用されるべきか否かを決定する。例えば、上述のように、文書の現在の状態iが画像を含まない場合、文書のこの状態へ再画像縮小や画像省略変形を適用する点がない。更に、画像の現在の状態iを得るために、「第1を除く全ての画像の省略」変形がすでに適用されている場合、この現在の状態iへ「第1を除く全ての画像の省略」変形を適用する点がない。

【0090】変形メモリ660で選択された現在の変形が文書の現在の状態iへ適切に適用できる場合、現在の状態iに対する変形された抽象構文ツリーによって指示されるように、変形メモリ660は、その状態に対する抽象構文ツリーにその変形を適用して子状態jを生成する。子状態jは、更に変形された抽象構文ツリー、及びこの子状態jに達するために必要な原文書から省略された内容に基づいて変形されたままのサブページを指示するサブページリストを含む。最後に、コントローラ610の制御下で、文書サイズ評価メモリ650は、その得られた文書が限られた表示領域装置510への出力のために十分に良いであるか否かを決定するために子状態j

で得られた文書の評価する。次に、その評価値は、新たに生成された子状態jに格納される。

【0091】変形メモリ660が新たな子状態jを生成した後、その状態jに対する変形抽象構文ツリーは、状態jに対応する文書のサイズ要求を評価するために、文書サイズ評価メモリ650に出力される。

【0092】変形済文書の第1のページに対する抽象構文ツリーが十分に良いと決定されると、その抽象構文ツリーは、ツリーから文書への再マッピングメモリ670へ出力され、そのメモリは、その抽象構文ツリーから第1の再オーサリング済サブページをレンダリングする。その第1の再オーサリング済サブページは、ツリーから文書への再マッピングメモリ670から入出力インターフェース620へ出力され、最後に、限られた領域表示装置520へ送信される。同時に、変形メモリ660は、原文書を第1の十分に良い再オーサリング済サブページへ変形することから得られるあらゆるサブページへ、追加の変形を適用し続ける。各そのようなサブページが十分に良いサブページへ変形されると、各そのような十分に良いサブページに対する抽象構文は、そのサブページに対する要求が限られた領域表示装置510から文書再オーサリングシステム600によって受信されるまで、再オーサリング済ページキャッシュ636に格納される。

【0093】そのサブページに対する要求が文書再オーサリングシステム600によって受信されると、そに対する要求されたサブページに対する抽象構文ツリーは、ツリーから文書への再マッピングメモリ670へ出力され、そのメモリは、その抽象構文ツリーから要求された再オーサリング済サブページをレンダリングする。そに対する要求された再オーサリング済サブページは、ツリーから文書への再マッピングメモリ670から入出力インターフェース620へ出力され、最終的に、限られた領域表示装置510へ送信される。

【0094】図6乃至図8に示されているメモリ及び他の構成要素の各々は、適切にプログラムされた汎用コンピュータの部分として実施され得ることを理解すべきである。或いは、図6乃至図8に示されているメモリの各々は、一つ又は複数のASIC内の論理的に個別のハードウェアメモリとして、FPGA、PLA又はPALを使用し、又は個別の論理エレメント又は別個のメモリ要素を使用して、実施され得る。図6乃至図8に示されているメモリの各々が取る特定のフォームは、設計上の選択事項であり、それらは、当業者にとって明で予測可能である。

【0095】リンク522、560及び580は、限られた表示領域装置510をホストノード570へ又はホストノード570を送受信装置通信システム550又は分散型ネットワークの残りの部分590へ接続のためにあらゆる既知の又は将来開発されるであろう装置又はシ

システムによって可能とされる。このように、リンク522、560及び580の各々は、直接ケーブル接続、ワイドエリアネットワーク又はローカルエリアネットワークを介する接続、イントラネットを介する接続、又はインターネットを介する接続として、実施され得る。一般に、リンク522、560及び580は、分散形ネットワークを介して対応する装置をホストノード570へ接続するために使用可能なあらゆる既知の又は将来開発される接続システム又は構造であり得る。

【0096】更に、文書再オーサリングシステム600は、プログラムされた汎用コンピュータ上で実施されることが好ましい事を理解すべきである。しかしながら、また、文書再オーサリングシステム600は、特殊目的のコンピュータ、周辺集積メモリ要素のようなプログラムされたマイクロプロセッサ又はマイクロコントローラ、及びASICや他の集積メモリ、デジタル信号プロセッサ、個別の要素メモリのようなハードワイヤードエレクトロニクス又は論理メモリ、PLD、PLA、FPGA、又はPAL、等で実施され得る。一般に、図1乃至図15に示されるフロー図を実行することが出来る有限状態マシンを実施出来るあらゆる装置は、文書再オーサリングシステム600を実施するために使用され得る。

【0097】図8に示されるメモリ630は、スタティック又はダイナミックRAMを使用して実施されるのが好ましい。しかしながら、メモリ630は、フロッピーディスク及びディスクドライブ、書き込み可能光学ディスク及びディスクドライブ、ハードドライブ、フラッシュメモリ又はあらゆる既知の又は将来開発される揮発性又は不揮発性変更可能メモリを使用して実施され得る。更に、メモリ630は、コントローラ610に対するコントロールプログラムを格納する一つ又は複数の部分を含み得る。一般に、そのようなコントロールプログラムは、フラッシュメモリ、ROM、PROM、及びEPROM又はEEPROM、CD-ROMとディスクドライブ、或いはあらゆる既知の又は将来開発される変更可能又は変更不能不揮発性メモリを使用して格納されることが好ましい。

【0098】図10は、他の例示の原文書及びその文書から生成される抽象構文ツリーの示す。図10に示されるように、文書は、画像、各2行3列の表、及びテキストを含む。このページから生成された得られた抽象構文ツリーは、“Page”（ページ）とラベル付けされた根ノードを含む。三つの中間ノード、即ち、画像、表及びテキストパラグラフの各々に対応する“Image”（画像）、“Table”（表）及び“Paragraph”（パラグラフ）は、夫々根“Page”ノードから延出する。更に、図10に示されるように、二つの行の各々に対応する、二つの中間ノード、即ち、“Row1”（行1）及び“Row2”（行2）は、夫々中間“Table”ノードから延出する。最後に、各行の三つのセルの各々に対応する三つのノードは、夫々“Row1”及び“Row2”ノードの各々から延出する。

【0099】例えば、図10に示されるページを再オーサリングするために、適用されるべき第1の変形は、一般に、完全サイズの画像を25%だけ縮小された画像を表すノードで置き換える。次に、完全サイズ画像に対応する根ノードを有する新たな抽象構文ツリーが形成され、変形された抽象構文ツリーの縮小画像ノードにハイパーテキストリンクによってリンクされる。25%縮小された画像を有する再オーサリング済ページがまだ十分に良くない場合、画像を50%、75%に縮小し及び次に画像を完全に除去する画像縮小変形は、十分に良い画像が得られるまで、原文書へ適用する。夫々の場合、抽象構文ツリーは、画像に対応する変形されたノードから完全サイズの画像を含む分岐の抽象構文ツリーへのリンクを含む。画像を完全に除去することでは、まだ十分に良い再オーサリング済文書を得るのに不十分である場合、上述のように、テーブルをセットのリンクされた個々のセルに変形するために表変形が適用できるか、或いはテキストパラグラフを別のサブページに移動するために第1センテンス省略変形が適用出来る。

【0100】図11及び図12は、本発明に従うページを再オーサリング擦るための例示の方法のアウトラインを示すフロー図である。図11及び図12に示されるように、制御は、ステップS100で開始してステップS110に続き、そこでユーザは、限られた表示領域を有する装置を本発明の再オーサリングシステムへ接続する。次に、ステップS120において、再オーサリングシステムは、限られた表示領域装置上への表示のために要求されたページを再オーサリング出来ることが必要な限られた表示領域についての必要な情報を得るために、一つ又は複数のパラメータフォームをユーザに送信する。次に、ステップS130において、再オーサリングシステムは、ユーザからパラメータ情報を入力し、メモリに入力されたパラメータ情報を格納する。次に、制御がステップS140に続く。

【0101】図6及び7に関連して上述されたように、ステップS120とS130でアウトラインで示されたパラメータ情報収集処理が自動化でき、それによって、ユーザは、ステップS120及びS130の実行に積極的に関与する必要はない。或いは、任意のステップS135で示されるように、ステップS120及びS130は、ステップS135によって置き換えられ得る。ステップS135において、この特定の限られた表示領域装置のための予め格納されたパラメータ情報を識別する識別コードを再オーサリングシステムへ、ユーザが積極的に入力するか、又は限られた表示領域装置が自動的に出力する。次に、制御がステップS140に続く。

【0102】ステップS140において、分散形ネットワーク上の文書要求は、限られた表示領域装置を使用して、ユーザから再オーサリングシステムへ出力される。次に、ステップS150において、再オーサリングシ

テムが分散型ネットワークから要求された文書を得る。次に、ステップS160において、得られた文書は、その文書の抽象構文ツリーを作るために解析される。次に、ステップS170において、得られた原文書のための評価値が抽象構文ツリーから生成される。次に、制御がステップS180に続く。

【0103】ステップS180において、評価値は、得られた文書が、何ら再オーサリングされることなく、限られた表示領域装置上に十分に良いに表示されるか否かを決定するために解析される。若しそうならば、制御がステップS340にジャンプする。そうでない場合、制御がステップS190に続く。

【0104】ステップS190において、一つ又は複数の予備再オーサリング変形は、得られた原文書の抽象構文ツリーに適用される。これらの予備再オーサリング変形は、例えば、内容は無いが表示領域を消費する原文書の部分を除去するために使用される。例えば、得られた文書のそのような部分は、バナー及び他のページやそのページの他の部分へのリンクを単に識別している他のグラフィカル要素を含む。これらの内容の無い画像は、テキストリンクによって置き換えられる。しかしながら、そのような変形は実際にその画像から何ら内容を除去しないので、このようにページを再オーサリングすることによって、除去された部分を保持する必要がある。原文書の内容に影響を及ぼすことなく、除去できる他の部分は、白スペース及び原文書に対する内容の無い審美的フォーマットリングを追加するフォーマットリングコマンドを含む。最後に、大きく複雑なフォントの不必要な表示領域要求を除去するために、文書の種々のフォントを単一の標準フォントへ変換する他の変形が適用できる。

【0105】ステップS190において、予備再オーサリング変形が適用されると、制御がステップS200に続き、そこで、予備再オーサリング済原文書に対する評価値が生成される。次に、ステップS210において、予備再オーサリング済文書の評価値は、予備再オーサリング済文書が限られた表示領域装置上に十分に良いに表示されるか否かを決定するためにチェックされる。そうでない場合は、制御がステップS220に続く。

【0106】ステップS220において、予備に再オーサリング済文書に対応する検索スペースの状態0が、検索スペースの現在の状態として選択される。次に、ステップS230において、第1の変形が現在の状態として選択される。次に、ステップS240において、現在の変形が現在の状態の抽象構文ツリーに適用出来るか否かを決定する。上記にアウトラインを示されたように、変形の種々のものは、その変形が効率的に現在の再オーサリング済文書に適用出来るか否か又は現在の変形が前に適用された変形と適切に組み合わせられるか否かを指示する条件を有する。現在の状態に対応する現在の再オーサリング済文書は、現在の変形が効率的に適用でき、前に

適用された変形とコンフリクトしない場合、制御がステップS250へ続く。そうでない場合、制御がステップS290へジャンプする。

【0107】ステップS250において、現在の状態が現在の変形を使用して子状態へ変形され、変形された抽象構文ツリー及びあらゆる結果のサブページを含む得られた子状態が検索スペースに追加される。次に、ステップS260において、評価値は、ステップS250において生成された子状態に対応する変形された抽象構文ツリーに対応する文書に対して生成される。次に、ステップS270において、その評価値は、ステップS250で生成された子状態に対応する文書が限られた表示領域装置上に十分に良いに表示されるか否かを決定するために解析される。その評価値が再オーサリング済文書又はサブページが十分に良いであることを指示すると、制御がステップS310にジャンプする。そうでない場合、制御がステップS280に続く。

【0108】ステップS280において、全ての状態が現在の状態へ適用されたか否かを決定する。変形の全てが適用されてはいない場合、制御がステップS290に続く。そうでない場合、制御がステップS300へジャンプする。

【0109】ステップS290において、次の変形が現在の状態として選択され、制御がステップS240にジャンプして戻る。反対に、ステップS300において、最良の評価値を有する検索スペースの状態が現在の状態として選択される。次に、制御がステップS230にジャンプして戻る。

【0110】ステップS310において、現在の状態によって定義された文書又はサブページが、要求される限られた表示領域装置への送信に連する最初に再オーサリング済ページ又は次の再オーサリング済サブページとして再オーサリング済ページキャッシュに追加される。次に、ステップS320において、再オーサリング済ページキャッシュへ追加された十分に良いサブページからのサブページがあるか否かを決定する。まだ再オーサリング待ちのそのようなサブページがある時、制御がステップS330に続く。そうでない場合、制御がステップS340にジャンプする。

【0111】ステップS330において、再オーサリング待ちサブページの一つに対応する検索スペースの状態が現在の状態として選択される。次に、制御がステップS230へジャンプして戻る。反対に、再オーサリング待ち更なるサブページが無いので、ステップS340において、第1の再オーサリング済ページが要求された限られた表示領域装置へ出力される。次に、ステップS350において、制御ルーチンが終了する。

【0112】図13は、本発明に従う省略変形の例示の実施の形態のアウトラインを示す。図13に示されるように、省略変形ルーチンはステップS40で開始して

ステップS410に続き、そこで除去されるべき現在のページ又はサブページの部分が選択される。次に、ステップS420において、選択された部分が新たなサブページにコピーされる。次に、ステップS430において、その選択された部分のための識別子が生成される。一般に、その識別子は、選択された部分のある内容を使用して生成される。例えば、選択された部分がパラグラフ又は他のテキストストリングである場合、その識別子は、選択された部分の最初のセンテンス又は最初のセンテンスの最初の部分である。選択された部分が画像の場合、その識別子は、ウェブ文書の画像を識別するために使用されるテキストの部分であり得る。次に、制御がステップS440に続く。

【0113】ステップS440において、リンクが現在のページ又はサブページを生成されたサブページにリンクするために生成される。次に、ステップS450において、選択された部分が現在のページ又はサブページから除去されて、その識別子とリンクが現在のページに追加される。次に、ステップS640において、制御ルーチンが終了する。

【0114】図14は、本発明に従う表変形の例示の実施形態のアウトラインを示す。図14に示されるように、表変形はステップS500で開始して、ステップS505に続き、そこでトップレベル表が現在の表として選択される。次に、ステップS510において、現在の表は、現在の表に入れ子表があるか否かを決定するためにチェックされる。もしそうならば、制御がステップS515に続き、そうでない場合、制御がステップS520にジャンプする。ステップS515において、現在の表の一つの入れ子表が新たな現在の表として選択される。次に、制御がステップS510にジャンプして戻り、現在の表として選択されたこの入れ子表に入れ子表があるか否かを決定する。

【0115】現在の表にもう入れ子表が無い場合、ステップS520において、現在の表は、現在の表にサイドバーがあるか否かを決定するためにチェックされる。もしそうならば、制御がステップS525に続く。そうでない場合、制御がステップS535へジャンプする。ステップS525において、リンクリストが、現在の表のサイドバーの全てのリンクの全てから生成する。次に、ステップS530において、リンクリストは、現在の表の最後に配される。次に、制御がステップS535に続く。

【0116】ステップS535において、現在の表が二つ又はそれより多い部分に分割される。特に、上述の方法は、表の各セルを別の部分に分割することである。次に、ステップS540において、現在の表の各部分は、別の新たなサブページにコピーし、"Next"及び"Previous"リンクが各サブページに追加される。次に、ステップ

S545において、現在の表は、ステップS540において生成されたリンクされたサブページのセットで置き換えられる。次に、制御がステップS550に続く。

【0117】ステップS550において、現在の表は、それがトップレベル表であるか否かを決定するためにチェックされる。もし否ならば、まだ複数の部分に分割されるべき少なくとも一つのより高いレベル表がある。従って、制御がステップS555に続く。そうでない場合、制御がステップS560にジャンプする。

【0118】ステップS555において、現在の表を含む表は、新たな現在の表として選択される。次に、制御がステップS510にジャンプして戻り、現在の表に入れ子表があるか否かを決定する。反対に、ステップS560において、制御ルーチンが終了する。

【0119】図15は、本発明に従う画像縮小変形の例示の実施形態のアウトラインを示す。ステップS600で開始し、画像縮小変形がステップS610に続き、現在のサブページで縮小されるべき画像が選択される。次に、ステップS620において、適用される特定の画像縮小変形に関連する縮小ファクタに基づいて、縮小された画像が生成される。次に、ステップS630において、現在のサブページは、選択された画像が前に縮小されているか否かを決定するために解析される。その場合、制御がステップS670にジャンプする。そうでない場合、制御がステップS640に続く。

【0120】ステップS640において、選択された画像が新たなサブページにコピーされる。次に、ステップS650において、この新たなサブページへのリンクが生成される。次に、ステップS660において、フルサイズ画像が現在のページ又はサブページから除去され、縮小された画像と生成されたリンクは、再オーサリング済ページを形成するために、現在のページへ追加される。次に、制御がステップS680にジャンプする。

【0121】反対に、ステップS670において、フルサイズの画像を現在のサブページから移動するのではなく、古い側の縮小された画像が、現在のサブページから除去され、その新たな縮小された画像が現在のサブページに追加される。しかしながら、現在のサブページがすでにフルサイズの画像を含む前に生成されたサブページに対するリンクを有するので、再び、現在のサブページへのリンクを追加したり、そのフルサイズの画像を格納する新たなサブページを生成する必要はない。次に、制御がステップS680に続き、そこで制御ルーチンが終了する。

【0122】文書の完全な自動再オーサリングを使用しても、セルラー電話で使用される非常に小さなテキストのみのタイプの表示に起因して、楽しく且つ有益な過去の上手なセルラー電話ウェブブラウジングを行うためには、一般的なウェブ文書においてしばしば単に情報が多すぎる。一般的に、これらの装置及びサービス

は、ユーザが特に求めている情報を検出し表示するために使用される。即ち、これらの装置及びサービスは、目標とする情報検索及び抽出のために使用される。本発明の文書フィルタリングシステム及び方法によって、ユーザは、構造ページナビゲーションコマンドを正則表現パターンマッチングとレポート生成関数（機能）とに組合せる簡単なエンドユーザスクリーニング言語を介して、彼等/彼女等が興味のある文書の部分のみを抽出することが可能となる。

【0123】1998年4月、オーストラリア、ブリスベン の第7回国際ワールドワイドウェブ会議でのR. MILLER 等による“SPHINX: a framework for creating personal, site-specific Web crawlers”に記述されているように、スフィンクス (SPHINX) システムは、本発明のシステム及び方法のフィルタリングメカニズムと機能性において類似するカスタム パーソナル ウェブクロウラをユーザに生成させる可視ツールを提供する。1998年4月、オーストラリア、ブリスベンの第7回国際ワールドワイドウェブ会議でのA. Sugiyama等による“Internet

Scrapbook: automating Web browsing tasks by programming-by-demonstration”に記述されているように、インターネットスクラップブック (Internet Scrapbook) によって、ユーザは、ウェブページからエレメント (要素) を可視的に選択して、次に、ウェブページが変化する“スクラップブック”中のこれらの要素を更新でき、本発明のシステム及び方法の特定のページに対するページエレメント検索と類似する機能を提供できる。また、幾つかの市販の製品は、例えば、コーポレートレポートリテラリングやデータベースビジュアル化のような他のアプリケーションのための類似の機能性を提供する。ランコム社 (Lanacom, Inc.) の HYPERLINK http://www.Headliner.com http://www.headliner.comに述べられている、ランコムのヘッドライナー (Lanacom's Headliner) 及びオン表示社 (OnDisplay Inc.) の HYPERLINK http://www.ondisplay.com http://www.ondisplay.com に述べられている、オン表示のセンターステージ (OnDisplay's CenterStage) の両方は、ユーザにウェブページの抽出する構造部分を指定させる可視エディターを提供する。しかしながら、これらのシステムの何れもユーザに正則表現又はキーワードに基づいて内容を抽出する能力を提供しない。

【0124】本発明の文書フィルタリングシステム及び方法は、ユーザによって高レベルスクリーニング言語で書かれたコマンドに基づいて、文書から部分情報を抽出する能力を有する。本発明の文書フィルタリングシステム及び方法は、上述の本発明の文書再オーサリングシステム及び方法を使用して抽出された情報の再オーサリングに加えて、ページ構造ナビゲーション、正則表現マッチング、サイトトラバース (走査)、即ち、ウェブクロウリング)、及び対話マッチングを組み合わせる。

【0125】フィルタスクリプトは、単にテキストファイルに入力され、ウェブサーバーにセーブされる。フィルタスクリプトは、ユーザがそのURLを要求する時は何時でも実行される。フィルタスクリプトは、一般的に、目標のウェブページをロードし、そのウェブページ内の（構造的に及び/又は正則表現によって記述される）特定の位置を走査し、これらの位置でその内容を抽出し、次に、抽出された内容を文書再オーサリングシステムを介してユーザに戻される前に適切にフォーマットされるために送出される。

【0126】本発明の文書フィルタリングシステム及び方法は、HTML文書内の“現在のコンテキスト”の概念を使用する簡単なセットのHTML文書ナビゲーションオブジェクトを提供することによって、本発明の文書再オーサリングシステム及び方法の解析ツリー生成及びナビゲーションの利点がある。現在のコンテキストは、HTML文書内の位置を参照する点において、データベースプログラミングの“カーソル”と類似する。

【0127】実際に、現在のコンテキストは、HTML解析ツリーのノードを参照する。ナビゲーションコマンドは、HTML文書の希望の部分が検出されるまで、この参照をツリー内で移動するように働く。この時に、希望の部分が抽出され得る。例えば、図10は、HTML文書及びその対応する解析ツリーを示す。“GO URL”コマンドを実行する事によって、文書が最初にロードされると、現在のコンテキストが解析ツリーの根ノードをポインティングしている。根ノード、本質的に、全体の文書を参照する。

【0128】図16は、ここでアウトラインが示された文書フィルタリングシステム及び方法を実施するフィルタメモリ690を更に含む文書再オーサリングシステム600の例示の実施の形態を示す。特に、コントロール610の制御下で、フィルタメモリ690は、ユーザによって通信リンク522又は560を介して要求されたフィルタを入力する。それに対する要求されたフィルタは、通信リンク580を介してそのようなフィルタを格納する分散形ネットワークのノードから供給される。次に、このフィルタメモリ690は、要求された文書に対して要求された文書を格納する分散形ネットワークのノードから入力し、それに対する要求された文書をフィルタリングして要求されたページの要素を抽出する。フィルタメモリ690は、最初に格納された原文書の代わりに、これらの抽出されたページ要素を原ページメモリ631に格納する。次に、文書再オーサリングシステム600は、これらの抽出されたページ要素上に、あたかもそれらが再オーサリング待ち原文書であるように、働く。

【0129】原文書からのページ要素の抽出において、フィルタメモリ690は、原文書から抽象構文ツリー生成メモリによって生成されると共に抽象構文ツリーメモ

り633に格納される抽象構文ツリーを使用する。

【0130】図17は、要求された文書がフィルタリングされるべき時、情報の流れの例示の実施の形態を示す。図17に示されるように、フィルタに対する要求が限られた表示領域装置510によってHTTPプロキシサーバ571へ出力された後、フィルタのそれに対する要求は、HTTPプロキシサーバ571によって、要求されたフィルタを格納する新たな分散形ネットワークの遠隔ノード592へ送られる。要求されたフィルタを格納する遠隔ノード592は、要求されたフィルタを文書フィルタ690へ戻す。次に、コントロール610の制御下で、文書フィルタ690は、要求されたページを格納する分散形ネットワークの遠隔ノード591から文書を要求する。要求されたページを格納する遠隔ノード591は、文書を文書フィルタ690に戻す。次に、文書フィルタ690は、遠隔ノード592から戻されたフィルタ及び抽象構文ツリー生成メモリ640によって生成された抽象構文ツリーを使用して、戻された文書をフィルタリングする。文書フィルタ690は、抽出されたページ要素を文書再オーサリングシステム600に戻し、そこで、抽出されたページ要素が、上述されたように、再オーサリングのための原文書として処理される。

【0131】ページナビゲーションコマンドは、次の三つのタイプがある。より指定的内容を選択するために現在のコンテキストに入るコマンド、現在のコンテキストから閉鎖構造へ出るコマンド、及び、例えば、現在のコンテキスト内に適切に含まれていても含まれていなくてもよいある種の次の構造にナビゲートするために、現在のコンテキストの開始からシーケンシャルにページを走査するコマンドである。

【0132】最も単純なタイプのナビゲーションコマンドは、現在のコンテキストに入る。例えば、図10に示される文書と現在のコンテキストが与えられると、コマンド"GO ROW 2"を実行すると、図18に示されるように、現在のコンテキストが現在のコンテキスト内で第2の表行オブジェクトへ移動される。

【0133】また、現在のコンテキストは、"GO ENCLOSING"コマンドを使用することによって、拡大される、即ち、解析ツリーを根ノードに向かって移動することが出来る。例えば、図18に示される文書とコンテキストが与えられると、"GO ENCLCING TABLE"コマンドによって図19に示される現在のコンテキストが得られる。

【0134】最後に、現在のコンテキストは、オブジェクトがユーザに見えるように、シーケンシャルにページ中のオブジェクト間で前後に移動され得る。これは、解析ツリーのプレフィックス走査内で現在のコンテキストをその現在の位置から前後に移動することによって完了される。これによって、最初に現在のコンテキスト内で実行され、次にそのページ上の現在のコンテキストに従うオブジェクトが続く検索を実行する。例えば、"GO P

REVIOUS IMAGE"コマンドが現在のコンテキストからシーケンシャルに検出される前の画像に移動する。

【0135】また、ネーム付けされたページ要素に加えて、正則表現を使用して、ナビゲーションコマンドが指定され得る。例えば、"GO NEXT""DOM\sJONES\s*(\d+)\s*sPOINTS""コマンドは、ページ上のテキストブロックのプレフィックス走査を使用して、現在のコンテキストを指定された正則表現の次のマッチングに移動する。本発明のフィルタリングシステム及び方法は、サブ表現を区分し、それらを出力ストリングに呼び出すことができる。

【0136】また、上述の単純なナビゲーションコマンドは、"LINKEDPAGE"ページオブジェクトタイプを使用して、一セットのリンクされたウェブページ間でナビゲートするために使用され得る。例えば、"GO FIRST LINKEDPAGE"コマンドは、現在のコンテキスト内で第1のハイパーテキストリンクへ移動し、参照されたページをロードし、現在のコンテキストをこの文書の解析ツリーの根へ移動すると共に、"GO ENCLCING LINKEDPAGE"コマンドは、現在のコンテキストを現在処理中の文書へのハイパーテキストリンクへ戻す。

【0137】ページ間の走査は、各々が特定のURLを有するスクリプト状態情報（現在のコンテキストを含む）と解析ツリーを対とする、スクリプトアクティベーションのスタックによって処理される。これは、リンクされたページ間で前後に迅速なナビゲーションを促進し、"GO ENCLCING LINKEDPAGE"コマンドを支援することを必要とする。

【0138】現在のコンテキストが、対象となるページオブジェクトへ移動されると、"REPORT"コマンドがそれを抽出するために使用される。この"REPORT"コマンドでフィルタスクリプト内で数回発行されることが出来る。その場合、抽出されたページ要素が連結される。また、"REPORT"コマンドは、任意のストリングを出力に挿入することが出来る。それは、正則表現パターンマッチングからのサブストリングを含むことが出来る。例えば、"REPORT" Dom:\1" コマンドは、ストリング" Dom: "+正則表現マッチングの間に抽出される識別子"1"によって識別されるサブストリングをフィルタ出力に追加する。

【0139】しばしば、ユーザは、ウェブページ上に特定の種類のページ要素がどのくらい多くあるかを予め知らない。例えば、日々の電子雑誌（eマガジン）中のニュース記事パラグラフの数を、一般に、予め知ることが出来ない。"FOREACH" コマンドは、指定された基準を満足する現在のコンテキスト内で検出された全てのページ要素に対して一連のコマンドを実行する事によって、この情報の欠落をアクセスする。"LINKEDPAGE" ターゲットが使用される時、これは、ウェブサイト内でリンクされたページの全てを訪問するウェブブラウザの機能を提供する。以上の例において、省略符号は、連続する有

効フィルタコマンドを表す。

【0140】「FOREACH PARAGRAPH」コマンドは、現在のコンテキスト(DO…END)内で各パラグラフに移動し、指定されたコマンドを実行する。

【0141】「FOREACH LINKEDPAGE」コマンドは、DO…END現在ページからハイパーテキストリンクを介して到達できる各ページをロードし、指定されたコマンドを実行する。

【0142】フィルタは、ナビゲーションの失敗、正則表現マッチング失敗又はウェブページ検索エラー等のあらゆる種類のエラーに遭遇すると何時でも、それは、オフエンディングコマンドが埋め込まれている最内「FOREACH」ループの次の反復を開始する。フィルタのトップレベルでエラーが生成すると、フィルタは、実行を中止し、未決出力を生成する。

【0143】本発明の文書再オーサリングシステム及び方法は、小さなスクリーンを有する表示上への表示のために文書の良い自動再オーサリングを行う。本発明の文書再オーサリングシステム及び方法の例示の実施の形態は、多数のスクリーンサイズに対する広範囲に亘るページに関して非公式にテストを行った。本発明の文書再オーサリングシステム及び方法のこの例示の実施の形態は、読み易くナビゲーション可能な出力を生成した。

【0144】例示の実施の形態において、本発明の文書再オーサリングシステム及び方法は、文書のためのスクリーン領域要求を推定するために、全ての画像とテキストのスペース要求を簡単に合計する。これは、ゼロックス年報(Xerox Annual Report)のような最小構造のかなり高密度の文書に対しては適切であるが、例えば、表のような、白スペースが多い又は先進のレイアウト技術を使用する文書にはうまく働かない。第2の例示の実施の形態において、本発明の文書再オーサリングシステム及び方法は、各文書バージョンを表示領域上にフォーマットする場合に、ブラウザによって実行される多くの作業を実行するサイズエスティミット(サイズ推定装置)を含む。ユーザが水平方向のスクロールを望まない場合、バンド幅、審美的基準のために、要求されたスクリーン領域以外のファクタ(例えば、再オーサリング済文書の実際の幅要求のような)が、含まれる必要があるかもしれない。

【0145】ユーザは、テキストを適合させるために、本発明の文書再オーサリングシステム及び方法に使用される種々のヒューリスティクスを調節できるべきである。例えば、ユーザは、変形技術の相対優先度を指定できるか、或いは全く使用されない幾つかの変形を指定できる。より高いレベルのアブストラクションにおいて、ユーザは、「より多くの内容」対「より大きな表示」のような、一セットのトレードオフに対するそれらの優先度を表すことが出来る。更に、本発明の再オーサリングシステム及び方法がクライアントに移動され、ブラウザ

と連結されることができ、それによって、ユーザは、ユーザがユーザの好みの結果を達成するまで、異なる変形をダイナミックに適用し、且つ取り消すことが出来る。

【0146】本発明の自動文書再オーサリングシステムと方法、特に、上述のHTTPプロキシサーバーの例示の実施の形態は、プログラムされた汎用コンピュータで実行されるのが好ましい。しかしながら、また、本発明の自動文書再オーサリングシステムと方法、特に、上述のHTTPプロキシサーバーは、特殊目的コンピュータ、マイクロコントローラと周辺集積メモリ要素、ASICや他の集積メモリ、デジタル信号プロセッサ、PLD、PLA、FPGA又はPALのような、ハードワイヤードエレクトロニクスや論理メモリ、等でも実行される。一般に、有限状態マシンを実施できるあらゆる装置は、本発明の自動文書再オーサリングシステムと方法、特に、上述のHTTPプロキシサーバーを実施するために使用できる。

【0147】本発明の自動文書再オーサリングシステム及び方法は、上述されたHTTPプロキシサーバーで実行される独立型再オーサリングプログラムを呼び出すことによって実行され得る。或いは、それは、ネットスケープナビゲーター(Netscape Navigator)等のような、従来のウェブブラウザへのプラグインを介して実行され得る。

【0148】更に、本発明の自動文書再オーサリングシステム及び方法は、ワールドワイドウェブから得られる文書の再オーサリングに関連して述べられたが、本発明の自動文書再オーサリングシステム及び方法は、ローカルエリアネットワーク、ワイドエリアネットワーク、イントラネット、インターネット、又はあらゆる他の分散形処理及び記憶ネットワークのような、あらゆる分散形ネットワークから得られる文書を再オーサリングするために使用され得る。

【0149】本発明は、上記にアウトラインが記された特定の実施の形態と共に記述したが、当業者には、多くの代替え、変更及びバリエーションは明白である。従って、上述の本発明の好適な実施の形態は、例示を意図しているが、本発明を限定する意図はない。種々の変化は、本発明の精神及び範囲を逸脱することなく、行われ得る。

【図面の簡単な説明】

【図1】本発明の文書再オーサリングシステム及び方法の一例示の実施の形態に従うセクションリストページ及び多数のセクションページへの文書の再オーサリングを示す図である。

【図2】本発明の文書再オーサリングシステム及び方法に従って複数のリンクされたセルに再オーサリングされるレイアウト表である。

【図3】本発明の文書再オーサリングシステム及び方法の一例示の実施の形態に従って異なる変形の適用に基づいて、文書がどのように異なる再オーサリング済状態に

再オーサリングされ得るかを示す図である。

【図4】本発明の文書再オーサリングシステム及び方法に従う表示情報をHTTPプロキシサーバへ供給するための制御フォームの例示の実施の形態を示す。

【図5】本発明の文書再オーサリングシステム及び方法に従う例示の文書の再オーサリングの一例示の実施の形態を示す。

【図6】本発明の文書再オーサリングシステム及び方法が使用される本発明の一例示の実施の形態のアウトラインを示すブロック図である。

【図7】本発明の文書再オーサリングシステム及び方法への文書フローの例示の実施の形態のアウトラインを示すブロック図である。

【図8】本発明の文書再オーサリングシステム及び方法の例示の実施の形態のアウトラインを示す機能ブロック図である。

【図9】本発明の文書再オーサリングシステム及び方法の文書バージョン検索スペースの例示の実施の形態の説明図である。

【図10】本発明に従う画像及びその画像から生成される抽象構文ツリーの例示の実施の形態の説明図である。

【図11】本発明に従う文書再オーサリングのための例

示の実施の形態のアウトラインを示すフロー図である。

【図12】本発明に従う文書再オーサリングのための例示の実施の形態のアウトラインを示すフロー図である。

【図13】本発明に従う省略変形を実行するための方法の例示の実施の形態のフロー図である。

【図14】本発明に従う表変形を実行するための方法の例示の実施の形態のフロー図である。

【図15】本発明に従う画像縮小変形を実行するための方法の実施の形態のフロー図である。

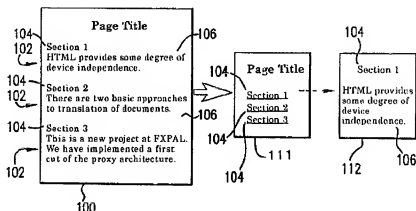
【図16】本発明に従う文書フィルタリングを含む本発明の文書再オーサリングシステム600の実施の形態のアウトラインを示す基本ブロック図である。

【図17】本発明に従う文書フィルタリング及び再オーサリングの間の文書フローの例示の実施の形態のブロック図である。

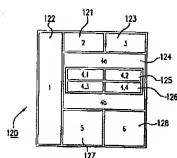
【図18】図10に示される画像から生成される抽象構文ツリー内をナビゲートするために本発明の文書フィルタリングシステム及び方法を使用する例示の実施の形態の説明図である。

【図19】本発明の文書フィルタリングシステム及び方法に従う図10の抽象構文ツリーの更なるナビゲートを示す説明図である。

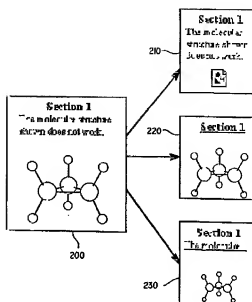
【図1】



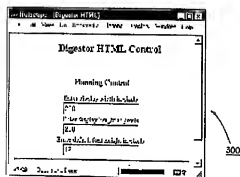
【図2】



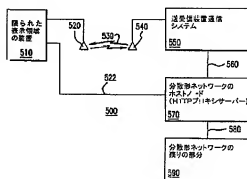
【図3】



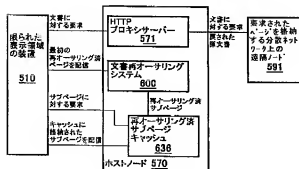
【図4】



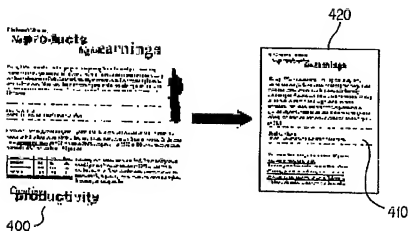
【図6】



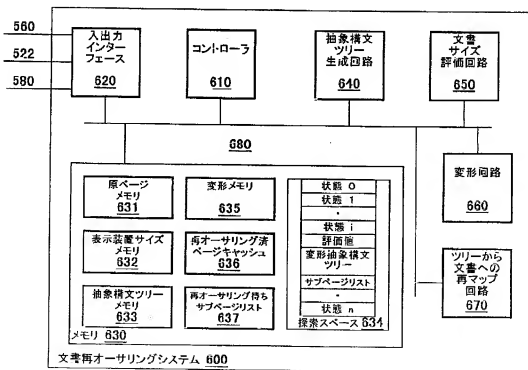
【図7】



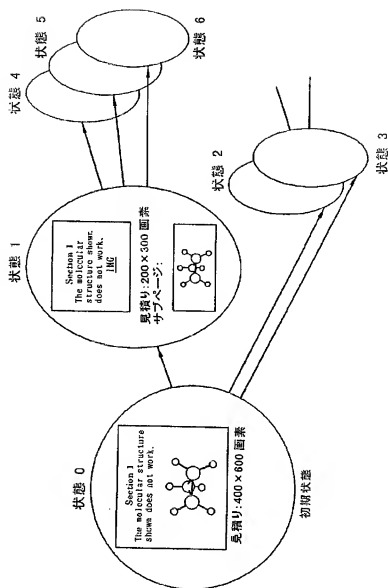
【図5】



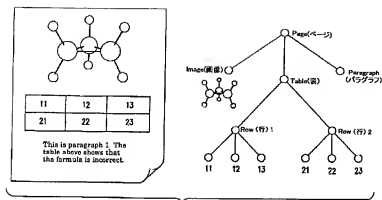
【図8】



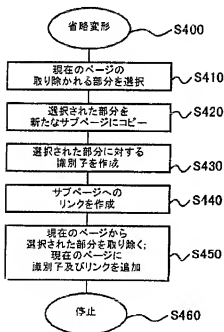
【図9】



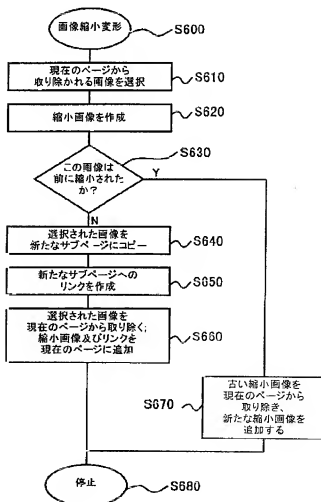
【図10】



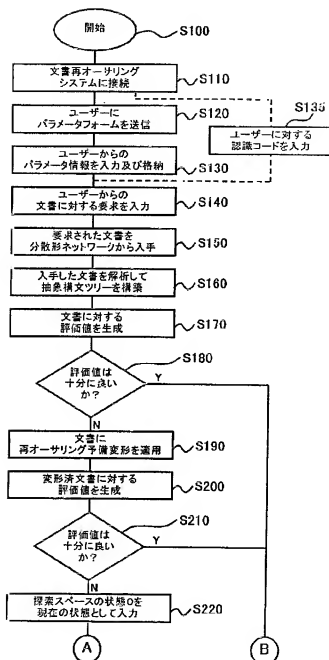
【図13】



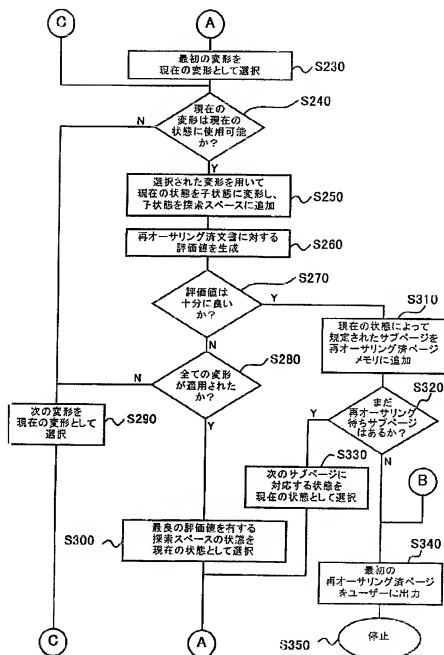
【図15】



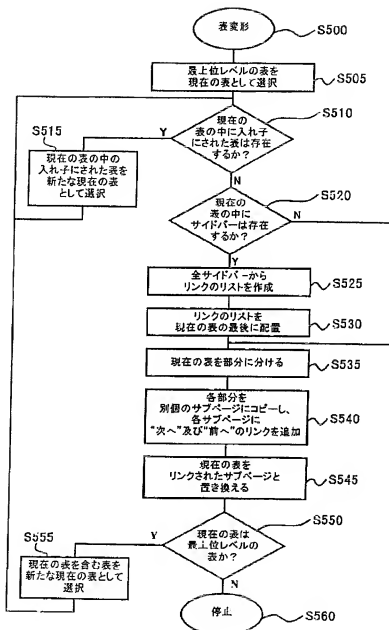
【図11】



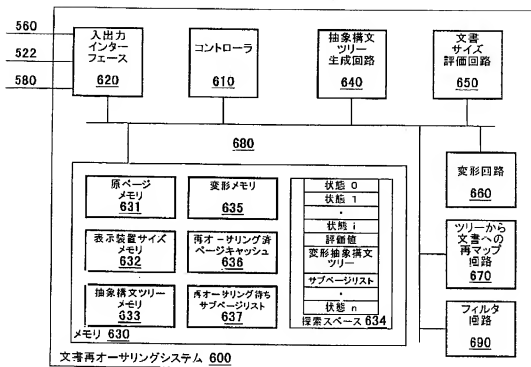
【図12】



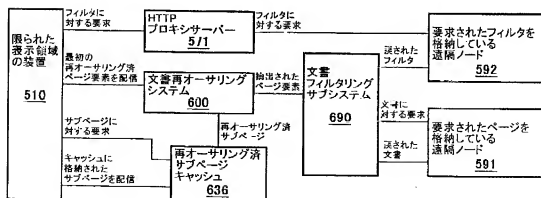
【図14】



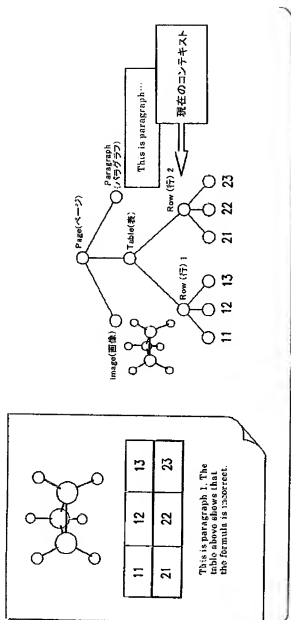
【図16】



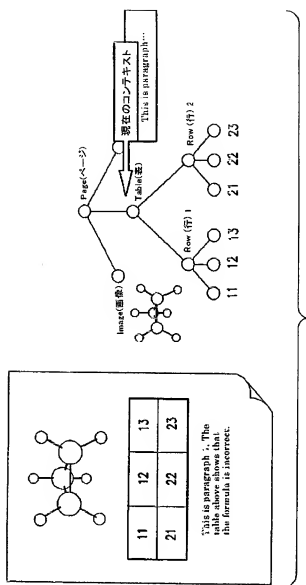
【図17】



【図18】



【図19】



フロントページの続き

(72)発明者 アンドレアス ガーゲンション
 アメリカ合衆国 94025 カリフォルニア
 州 メンロ パーク ウェイヴァリー ス
 トリート 210 ナンバー4

(72)発明者 ジョセフ ダブリュー、スリーヴァン
 アメリカ合衆国 94107 カリフォルニア
 州 サンフランシスコ ブルックサム ス
 トリート 175 ナンバー103